EFFICIENT CHANGE DETECTION METHODS FOR BIO AND HEALTHCARE SURVEILLANCE

A Thesis
Presented to
The Academic Faculty

by

Sung Won Han

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2010

EFFICIENT CHANGE DETECTION METHODS FOR BIO AND HEALTHCARE SURVEILLANCE

Approved by:

Professor Kwok-Leung Tsui, Advisor,
Committee Chair
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Yajun Mei, Advisor
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor David M. Goldsman
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Roshan Joseph Vengazhiyil
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor William H. Woodall
Department of Statistics
Virginia Tech

Date Approved: June 7th, 2010

To my mother, sister, and wife.

ACKNOWLEDGEMENTS

I would like to thank professors and students for their support and assistance. They helped me complete this dissertation. I express my sincere gratitude to my advisors, Kwok Tsui and Yajun Mei. Kwok Tsui shared with me his broad knowledge, insight, and wonderful guidance in the statistical detection methodologies. His encouragement made my Ph.D. study very enjoyable. Yajun Mei offered me a theoretical foundation in tackling my thesis problems. His critical comments made this thesis more rigorous. My thanks are also extended to the other committee members, David Goldsman, Roshan Joseph Vengazhiyil, and William H. Woodall for their valuable comments on this dissertation and to Jane Chisholm and Hae Young Kim for my thesis editing.

I also thank my friends, staff, and other professors in the School of Industrial and Systems Engineering and the School of Mathematics. Finally, I thank my mother and sister for their endless support. The special thanks goes to my wife, Hae Sun Kim, for her support of my study. Completing my Ph.D. program would have been impossible without their care and encouragement.

TABLE OF CONTENTS

DEDICATION iii

ACKNOWLEDGEMENTS iv

LIST OF TABLES viii

LIST OF FIGURES ix

SUMMARY xi

I INTRODUCTION 1

1.1 Contributions 2

1.2 Outline of the Thesis 3

II BACKGROUND 5

2.1 The Sequential Change Point Detection Problem 5

2.2 Performance Measures 6

2.3 The Detection Schemes and Literature for Temporal Surveillance 7

2.3.1 The CUSUM Procedure 7

2.3.2 The Shiryaev-Roberts Procedure 8

2.3.3 The EWMA Procedure 8

2.3.4 The Scan Statistics 9

2.4 Our Data Sets 10

2.4.1 Male Thyroid Cancer in New Mexico 10

2.4.2 Female Breast Cancer in New Hempshire 10

III A COMPARISON OF CUSUM, EWMA, AND SCAN STATISTICS FOR THE DETECTION OF RATE INCREASES IN HOMOGENEOUS POISSON DATA 14

3.1 Introduction 14

3.2 Problem Formulation 15

3.3 Detection Methods 16

3.3.1 Scan Statistics 16

3.3.2 Cumulative Sum Charts 17

3.3.3 Exponential Weighted Moving Average Charts 18
3.4 Simulation Study 19
3.4.1 Parameter Selection for Target ARL_0 19
3.4.2 Comparison of $CED(\nu, \lambda_1)$s at Different Points of Time for Changes ν at Fixed Shift Size λ_1 20
3.4.3 Comparison of $CED(\nu, \lambda_1)$s Under Different Shift Sizes 21
3.5 Example: The Detection of Increased Incidence in Male Thyroid Cancer 22
3.6 Conclusions 23
IV EARLY DETECTION OF RATE INCREASES IN NON-HOMOGENEOUS POISSON DATA 30
4.1 Introduction 30
4.2 Mathematical Formulation 31
4.3 The GLR Scheme and its Asymptotic Optimality Properties 33
4.4 Two Alternative Methods 35
4.5 Example Revisited 36
4.5.1 Model for Population Growth 37
4.5.2 Parameters in the Change-point Problem and Detection Schemes 38
4.5.3 When to Raise an Alarm 39
4.6 More Simulation Study 40
4.7 New Asymptotic Analysis 42
4.7.1 Asymptotic Analysis 44
4.7.2 Numerical Simulations 48
4.8 Conclusion 50
4.9 Appendix: Proof of Mathematical Results 50
4.9.1 Proof of Theorem 4.1 50
4.9.2 Proof of Theorem 4.2 52
4.9.3 Proof of Theorem 4.3 55
4.9.4 Proof of Theorem 4.4 56
4.9.5 Proof of Theorem 4.5 57

V SPATIOTEMPORAL SURVEILLANCE BASED ON LIKELIHOOD RATIO . 65
5.1 Introduction . . . 66
5.2 Notation for Positions, Coverage, and Statistics . . . 66
5.3 Spatiotemporal Surveillance Based on LR Statistics . . . 67
5.3.1 Problem Formulation . . . 68
5.3.2 Surveillance Methods Based on Likelihood Ratios . . . 69
5.3.3 Performance Comparison . . . 71
5.4 New Mexico Thyroid Cancer Example . . . 74
5.5 Conclusion . . . 77
VI SPATIOTEMPORAL SURVEILLANCE METHODS IN THE PRESENCE OF SPATIAL CORRELATION . . . 82
6.1 Introduction . . . 84
6.2 MCUSUM Methods for Spatiotemporal Surveillance . . . 85
6.2.1 Problem Definition . . . 85
6.2.2 When the Change Position and Time Are Unknown . . . 86
6.2.3 When the Change Position, Coverage and Magnitude Are Unknown . . . 88
6.3 Performance Analysis . . . 88
6.4 Example - Breast Cancer in New Hampshire . . . 93
6.5 Conclusions . . . 95
VITA . . . 111

LIST OF TABLES

1 A set of the parameters of the scan statistic method with different shift sizes given the targeted ARL_0=1500 25

2 A set of the parameters of the CUSUM charts with different shift sizes given the targeted ARL_0=1500 25

3 A set of the parameters in the EWMA charts with different shift sizes given the targeted ARL_0=1500 25

4 Estimated parameters for the population growth model in (29) 37

5 Detection delays of the three proposed detection schemes at change-point $\nu = 1$ or ω 58

6 ARL of the scan statistics when the change position and time are unknown (the change magnitude is 0.5) 73

7 ARL of the scan statistics when the change position, radius, and time are all unknown (the change magnitude is known to be 0.5) 74

8 Comparison of ARL for Different Surveillance Methods 92

LIST OF FIGURES

1 The three time series data related to male thyroid cancer in New Mexico during 1973-2005 illustrate the following: the total number of male thyroid cancers over years (the top panel), the trend of male population (the middle panel), and the crude incidence per 100,000 population (the bottom panel). . 12

2 Mortality per 10,000 for female breast cancer in New Hampshire 13

3 Comparison of the three detection methods with optimal parameters for given shift size λ_1 across different points of time for change ν. The x-axis represents time point of rate change (νs) and the y-axis represents $CED^u(\nu, \lambda_1)$s. 26

4 Comparison of the three detection methods with optimal parameters for a target shift size λ_1^* (ν=1). Sets of parameters for the scan statistics, CUSUM, and EWMA for a target size λ_1^* and change times ν=1 were used. The x-axis represents a true shift size (λ_1) and the y-axis represents $\frac{CED(\nu=1,\lambda_1)}{CED^{optimal.CUSUM}(\nu=1,\lambda_1)}$. 27

5 Comparison of the three detection methods with optimal parameters for a target shift size λ_1^* (ν=1,500). Sets of parameters for the scan statistics, CUSUM, and EWMA for a target size λ_1^* and change times ν=1,500 were used. The x-axis represents a true shift size (λ_1) and the y-axis corresponds to $\frac{CED(\nu=1,500,\lambda_1)}{CED^{optimal.CUSUM}(\nu=1,500,\lambda_1)}$. 28

6 The trend of male thyroid cancer incidence between 1973 and 2005. 29

7 Plots of statistics of the scan statistic method with m=37, the CUSUM chart with λ_0^*=2 and λ_1^*=2.5, the EWMA chart with α=0.02. The circle indicates the first time point when each method triggers an alarm. 29

8 Population and estimate in New Mexico during 1973-2005. The plot shows the actual observed and model-estimated male population sizes in New Mexico during 1973-2005. 38

9 The *left* panel plots the GLR statistics W_n over time n, as well as the alarm boundaries of $T_{GLR}(a)$ (solid line) and $T_{ATM}(c)$ (the dotted line). The *right* panel plots the WLR statistic $\hat{W}_n$ as well as the boundary of $T_{WLR}(b)$. . . . 38

10 The population sizes are from the smooth model in (29). The *top left* panel plots three different population size curves that correspond to the three cases considered. The other three panels illustrate the detection delays of the three proposed detection schemes as a function of change-point ν under different cases of the population size models. The plots suggest that the WLR scheme $T_{WLR}(b)$ seems to be the best if the population sizes are increasing (Cases A and B), whereas the GLR scheme $T_{GLR}(a)$ seems to be the best if the population sizes are decreasing (Case C). Moreover, the ATM scheme $T_{ATM}(c)$ seems to be *robust* under Lorden's worst-case detection delay criterion no matter whether the population sizes are increasing or decreasing. 63

11 The detection delays of the three proposed detection schemes at different change-points ν when the population sizes are given by the step functions. Left Panel: the step function is increasing. Right Panel: the step function is decreasing. 64

12 Outbreak patterns in the simulation 78

13 Vector and coordinate expressions of regional data (p=36) 78

14 Outbreak position $c = (i, j)$, radius r, and coverage A_c^r 78

15 The map of counties in New Mexico and the X-Y coordinate representation 78

16 Male thyroid cancer incidence per 100,000 in New Mexico 79

17 Scaled scan statistics for male thyroid cancer data in New Mexico 80

18 Clusters detected for the male thyroid cancer data in r=5 cases 81

19 Clusters detected for the male thyroid cancer data in variable radius cases . 81

20 Three symmetric outbreak patterns (S-1, S-5, and S-13) 97

21 The ARL comparison of Rac-MCUSUM and Max-MCUSUM methods when correlations are present (outbreak S-1 and $\delta = 1$) 97

22 The impact of spatial correlations on different surveillance methods ($\rho = 0.5$ and $\delta = 1$) (solid line: Max-MCUSUM, dashed line: Rac-MCUSUM with a known magnitude) 98

23 The map of New Hampshire and the X-Y coordinate 99

24 Surveillance statistics with $r = 0$ for breast cancer data in New Hampshire . 99

25 Surveillance statistics with $r = 5$ for breast cancer data in New Hampshire . 100

26 The cluster detected by the methods 100

SUMMARY

For the last several decades, sequential change point problems have been studied in both the theoretical area (sequential analysis) and the application area (industrial SPC). In the conventional application, the baseline process is assumed to be stationary, and the shift pattern is a step function that is sustained after the shift. However, in biosurveillance, the underlying assumptions of problems are more complicated. This thesis investigates several issues in biosurveillance such as non-homogeneous populations, spatiotemporal surveillance methods, and correlated structures in regional data.

The first part of the thesis discusses popular surveillance methods in sequential change point problems and off-line problems based on count data. For sequential change point problems, the CUSUM and the EWMA have been used in healthcare and public health surveillance to detect increases in the rates of diseases or symptoms. On the other hand, for off-line problems, scan statistics are widely used. In this chapter, we link the method for off-line problems to those for sequential change point problems. We investigate three methods–the CUSUM, the EWMA, and the scan statistics–and compare them by conditional expected delay (CED).

The second part of the thesis pertains to the on-line monitoring problem of detecting a change in the mean of Poisson count data with non-homogeneous population sizes. The most common detection schemes are based on generalized likelihood ratio statistics, known as an optimal method for the i.i.d. models. We propose alternative detection schemes based on the weighted likelihood ratios and the adaptive threshold method, which perform better than generalized likelihood ratio statistics in an increasing population. The properties of these three detection schemes are investigated by both theoretical analysis and numerical

simulation.

The third part of the thesis investigates spatiotemporal surveillance based on likelihood ratios. This chapter proposes a general framework for spatiotemporal surveillance based on likelihood ratio statistics over time windows. We show that the CUSUM and other popular likelihood ratio statistics are the special cases under such a general framework. We compare the efficiency of these surveillance methods in spatiotemporal cases for detecting clusters of incidence using both Monte Carlo simulations and a real example.

The fourth part proposes multivariate surveillance methods based on likelihood ratio tests in the presence of spatial correlations. By taking advantage of spatial correlations, the proposed methods can perform better than existing surveillance methods by providing the faster and more accurate detection. We illustrate the application of these methods with a breast cancer case in New Hampshire when observations are spatially correlated.

CHAPTER I

INTRODUCTION

In health care and bio surveillance, the timely detection of an increase in the rate of a disease is crucial because of the emergency situations resulting from outbreaks. For example, during the 2002-2003 Severe Acute Respiratory Syndrome (SARS) epidemic in China, approximately 650 people died. Another example is N1H1; a severe epidemic occurred in the United States and Mexico last year. If we had been able to detect the spread of such diseases earlier and identify the epicenter of the disease accurately, then we could have taken proper public health measures to reduce the number of victims. However, the problem in health care surveillance is challenging, so further statistical research is needed. Statistical challenges in health care and bio surveillance are summarized in Shmueli *et al.* [92]. More literature in health care surveillance is in Fienberg *et al.* [22], Woodall [110], and Tsui *et al.* [102].

One statistical methodology for dealing with healthcare surveillance is from the sequential change point detection problem. In the classical sequential change detection problem, one monitors the sequence of observations Y_1, Y_2,... from the process of a system. Initially, the process is under the *in-control* state (normal state). At some unknown time ν, the process goes to an *out-of-control* state (abnormal state). The problem is to detect a change in the process state quickly while controlling the false alarm rate. Motivated by engineering applications, the classical problem assumes that observations are independent and that the distributions in the process are known. The problem has been solved using several popular methods including Shewhart's chart, Roberts' EWMA, Page's CUSUM, and the Shiryaev-Roberts procedure. Some pioneering foundational work in the field includes Page [70], Shiryaev [91], Roberts [83], Lorden [57], Pollak [75], Moustakides [67], Ritov [81], Yakir

[113], and Lai [54]. For the more recent reviews, we refer to Basseville *et al.* [7], Lai [54], Peskir *et al.* [71], Poor *et al.* [77], and Woodall [110], and the references therein.

Unfortunately, the biosurveillance problem is more challenging than the classical change point detection problem or industrial SPC. It often assumes that the baseline is neither stationary nor independent, but it might exhibit a systematic pattern from seasonal trends or population changes. Furthermore, spatiotemporal biosurveillance should consider spatial features, and regional data might be correlated. This non-traditional problem assumption provides abundant new research opportunities in finding efficient detection schemes for the sequential change point problem. This thesis discusses some issues in biosurveillance such as non-homogeneous population baseline and regional correlations, and it proposes new detection schemes based on the context of the sequential change point detection problem.

1.1 Contributions

In Chapter 3, we compare the well-known detection methods–the CUSUM, the EMWA, and the scan statistics–for the detection of rate increase in homogeneous Poisson data. We also link the method for the off-line problem to methods for the sequential change point problem. We show how scan statistics in off-line problems are related to the moving average methods in the sequential change detection problem and how the CUSUM is related to the scan statistics with a variable time window in the off-line problem. In a simulation study, we show that Poisson CUSUM or EWMA procedures usually outperform the Poisson scan statistics. A comparison of the CUSUM and the EWMA shows that the CUSUM performs better under a larger shift with a later change time. However, the EWMA was superior to the CUSUMs in dealing with a smaller shift and an earlier change time.

In Chapter 4, motivated by the early detection of an increase in disease rates in non-homogeneous Poisson data, we investigate the performance of the generalized likelihood ratio statistics scheme, and develop new detection schemes. The generalized likelihood

ratio statistics scheme is an optimal method under Lorden's criteria for the i.i.d. models. However, our numerical simulation illustrate that the weighted likelihood ratios and the adaptive threshold methods we propose perform better than the generalized likelihood ratio statistics scheme if population sizes are increasing. To explain this, we also derive asymptotic properties of the worst detection delays of the three detection schemes with non-homogeneous population sizes.

In Chapter 5, we investigate existing and new spatiotemporal surveillance methods based on likelihood ratios. This chapter proposes a general framework for spatiotemporal surveillance based on likelihood ratio statistics over time windows. We show that CUSUM procedures are the special cases under the general framework. We compare the performance of these surveillance methods in spatiotemporal cases to detect outbreak clusters using Monte Carlo simulations and a real example.

In Chapter 6, we propose multivariate surveillance methods generalized from detection methods in multivariate control charts based on likelihood ratio tests in the presence of spatial correlations. We compare them using Monte Carlo simulation and show that the proposed methods perform better than existing surveillance methods and detect outbreaks more efficiently under the spatial correlation. We illustrate the application of these methods with a breast cancer example in New Hampshire when observations are spatially correlated.

1.2 Outline of the Thesis

In Chapter 2, we present an overview of the sequential change detection problem. We summarize the problem formula, performance measures, and test statistics. We also describe the data we used and review the literature.

Chapter 3 reviews the scan statistic, the CUSUM, and the EWMA procedures for Poisson distribution. We compare the performance of the detection methods by simulation under various scenarios. We also present a real case study for the detection of rate increases in male thyroid cancer in New Mexico. This chapter is an excerpt from Han, Tsui,

Ariyajunyab, and Kim [32], published in *Quality and Reliability Engineering International*, 2010, Vol. 26, Pages 279–289, and the excerpt format is based on the Copyright Transfer Agreement (2009) in Wiley Blackwell.

Chapter 4 states the mathematical formulation of the change point detection problem in a non-homogeneous population and explains the generalized likelihood ratio (GLR)-based scheme and its asymptotic optimality properties. We also propose two alternative schemes utilizing the effect of population size. Then, we apply the detection methods to the male thyroid cancer data and simulation study. Next, we introduce an asymptotic optimality theory in a new problem setting and investigate the asymptotic properties of the three proposed methods to gain better insights into the results of the finite-sample simulation. This chapter is an extended version of Mei, Han, and Tsui [63], which was accepted by *Statistica Sinica* for publication.

Chapter 5 introduces the general framework of likelihood ratio test statistics. Then we discuss likelihood ratio methods for spatiotemporal surveillance under the independence assumption. We also present a real application of male thyroid cancer data in New Mexico. This chapter is the part of Tsui, Han, Jiang, and Woodall [103] that has been submitted to *Technometrics*.

Chapter 6 presents likelihood-ratio-based methods for spatiotemporal surveillance with spatially correlated data. We also apply them to breast cancer data in New Hampshire to investigate the performance of the surveillance methods when spatial correlations exist. This chapter is an excerpt from Jiang, Han, Tsui, and Woodall [43], which will be published in *Statistics in Medicine*, DOI: 10.1002/sim.3877, and the excerpt format is based on the Copyright Transfer Agreement (2009) in Wiley Blackwell.

CHAPTER II

BACKGROUND

In this chapter, we discuss the general setting of sequential change point detection problems as well as well-known detection methods such as the CUSUM, the Shiryaev-Roberts, the EWMA, and the scan statistic. For a more detailed background, we refer readers to Basseville *et al.* [7].

2.1 The Sequential Change Point Detection Problem

Assume that a sequence of independent random variables $\{Y_1, Y_2, \ldots\}$ is observed over time. Initially, the process is under the in-control state (normal state) in the sense that the true distribution of Y_i is $f(\cdot)$. At an *unknown* time ν when an undesired event occurs, the process goes to the out-of-control state (abnormal state), and the distribution changes from $f(\cdot)$ to $g(\cdot)$. In other words, for some $\nu \geq 1$ (possibly ∞), $Y_1, \ldots, Y_{\nu-1}$ are distributed according to $f(\cdot)$, and $Y_\nu, Y_{\nu+1}, \ldots$ are distributed according to $g(\cdot)$. The objective is to detect such a change as quickly as possible while controlling the false alarm rate.

This problem is sequential change-point detection problem, and it can be formulated by sequentially testing the simple null hypothesis

$$H: \; Y_i \sim f(\cdot) \quad \text{for all } i \geq 1 \text{ (i.e., no change)}$$

against the composite alternative hypothesis

$$K: \begin{cases} Y_i \sim f(\cdot), & \text{if } 1 \leq i \leq \nu - 1 \\ Y_i \sim g(\cdot), & \text{if } i \geq \nu \end{cases} \quad \text{for some } \textit{unknown } \nu = 1, 2, 3, \ldots \text{ (i.e., a change occurs).}$$

In the sequential change point detection problem, a statistical procedure u can often be defined as a stopping time $T^u =$ first time n such that a detection statistic $U_n \geq h$ for a

pre-determined threshold h, where the detection statistic U_n at time n depends only on the first n observations $Y_1, \ldots, Y_n$. When stopping time $T^u = n$, the corresponding statistical procedure will raise an alarm at time n to indicate that a change occurs somewhere in the first n observations.

2.2 Performance Measures

To evaluate the performance of a statistical procedure u in the sequential change-point detection problem, one needs to define two criteria: the *false alarm* under the in-control state and the *detection delay* under the out-of-control state. In the literature, the false alarm criterion is generally evaluated by $E[T^u|\nu = \infty]$, which is also called *average run length* under the in-control state (ARL_0^u). Here, the change time $\nu = \infty$ indicates that there are no changes. Intuitively, the detection delay is $T^u - \nu$ when $T^u > \nu$. However, the rigorous definition needs to take into account the fact that T^u is a random variable. To deal with the randomness of T^u, a standard idea is to consider the *conditional expected delay* defined by $CED^u(\nu) = E[T^u - \nu|T^u \geq \nu]$ or the *worst detection delay* defined by $esssup E_\nu((T-\nu)^+|Y_1, ..., Y_{\nu-1}, T \geq \nu)$. Then, given a true alternative g, one would like to find a detection method u that minimizes $CED^u(\nu)$, subject to

$$ARL_0^u \geq \gamma, \tag{1}$$

where $\gamma(> 0)$ is the prespecified lower bound.

Unfortunately, the change time ν is also unknown, and thus $CED^u(\nu)$ cannot be used directly in the minimax formulation. Several approaches have been proposed. Shiryaev [91] and Pollak [75] proposed $\sup_{1\leq\nu<\infty} CED^u(\nu)$, which takes a supreme of CED for all time points. Another widely-used measure is $CED^u(\nu = 1)$, called the *initial-state average run length* under the out-of-control state (ARL_1), which is the detection delay when a change occurs at time $\nu = 1$. Third, Srivastava *et al.* [96] proposed $\lim_{\nu\longrightarrow\infty} CED^u(\nu)$, which is called the steady-state or stationary ARL under the out-of-control state. In many detection schemes such as the CUSUM, $\sup_{1\leq\nu<\infty} CED^u(\nu) = CED^u(\nu = 1)$ because

the initial-state average run length is usually the worst conditional expected delay. For the Shewhart chart, all three measures are equivalent since it utilizes the information of only the most current observation.

Based on the measure $esssupE_\nu((T-\nu)^+|Y_1,...,Y_{\nu-1},T\geq\nu)$, Lorden [57] proposed $\bar{E}_1(T) = \sup_{1\leq\nu\leq\infty} esssupE_\nu((T-\nu)^+|Y_1,...,Y_{\nu-1},T\geq\nu)$. It is closely related and asymptotically equivalent to the conditional expected delay: $\sup_{1\leq\nu<\infty} CED^u(\nu,\lambda_1)$.

2.3 The Detection Schemes and Literature for Temporal Surveillance

In this subsection, we discuss the following well-known detection methods: the CUSUM procedure, Shiryaev-Roberts procedure, EWMA procedure, and scan statistics. For the comprehensive review, we refer readers to Montgomery [64] and Woodall [109].

2.3.1 The CUSUM Procedure

The CUSUM procedure (Page [70]) is defined by $T^C = \inf\{n\geq 1|C_n\geq h\}$, where h is a threshold and the statistic C_n is the CUSUM statistic defined by

$$C_n = \max_{1\leq\nu\leq n}\{\prod_{j=\nu}^{n}\frac{g(Y_j)}{f(Y_j)}\}, \tag{2}$$

which is the maximum likelihood ratio over all possible time windows. The optimality of the CUSUM under the independent and identical baseline was established in Lorden [57], Moustakides [67], and Ritov [81]. The CUSUM procedure T^C can be rewritten as $T^{C'} = \inf\{n\geq 1 : C_n'\geq \log h\}$, where $C_n' = \max\{0,\log C_n\}$. In practice, the CUSUM procedure is widely used partly because C_n' has a recursive form:

$$C_n' = \max(0, C_{n-1}' + \log\frac{g(Y_n)}{f(Y_n)}). \tag{3}$$

In health care surveillance, Hill $et\ al.$ [37] and Weatherall $et\ al.$ [106] applied the CUSUM procedure to the surveillance of congenital malformations. Brook *et al.* [12], Hawkins *et al.* [35], and Yashchin [115] discussed the detection of a mean change in the Poisson distribution by the CUSUM procedure. Lucas [59] discussed the average run length

of the CUSUM charts under Poisson data. White $et\ al.$ [107] investigated a threshold for obtaining a target ARL_0 by Markov chain approximation. The comprehensive review for the CUSUM procedure under other distributions is in Hawkins *et al.* [35].

2.3.2 The Shiryaev-Roberts Procedure

The Shiryaev-Roberts procedure (Shiryaev [91] and Roberts [83]) is defined by $T^S = \inf\{n \geq 1 | S_n \geq h\}$, where S_n is the Shiryaev-Roberts statistic

$$S_n = \sum_{\nu=1}^{n} \{ \prod_{j=\nu}^{n} \frac{g(Y_j)}{f(Y_j)} \} \tag{4}$$

and h is a threshold. The recursive form of S_n is

$$S_n = \frac{g(Y_n)}{f(Y_n)}[S_{n-1} + 1].$$

This procedure calculates the summation of all possible likelihood ratios, which is based on the Bayesian motivation that all possible time windows have an equal chance for change (Shiryaev [91]). A theoretical advantage of this procedure is that $S_n - n$ is a martingale under P_∞. The performance of the Shiryaev-Roberts procedure is very close to the CUSUM procedure, see Pollak and Siegmund [75].

2.3.3 The EWMA Procedure

The EWMA procedure (Roberts [82]) is defined by $T^E = \inf\{n \geq 1 | E_n \geq h\}$, where

$$E_n = \alpha Y_n + (1 - \alpha) E_{n-1}, \tag{5}$$

where α is a weight parameter ($0 < \alpha \leq 1$) and $E_0 = E[Y]$. The EWMA statistic can be applied to log likelihood ratios, defined by

$$E_n = \alpha \log \frac{g(Y_n)}{f(Y_n)} + (1 - \alpha) E_{n-1}. \tag{6}$$

The EWMA statistic (6) can be simplified to the one based on the actual observation (5) if the distribution $f(\cdot)$ and $g(\cdot)$ belong to the same exponential family (Basseville *et al.* [7]).

The EWMA has been studied by Hunter [39], Crowder [17] [18], Ng *et al.* [69], and Lucas *et al.* [60], which developed the EWMA procedure for two-sided tests. Robinson *et al.* [84] and Shu *et al.* [93] discussed a one-sided EWMA procedure. Gan [27] modified the EWMA procedure to monitor the mean shift in the Poisson distribution. Borror *et al.* [11] [10] investigated the ARL of the Poisson EWMA procedure by a Markov chain simulation.

Many studies compared the performance of the CUSUM and EWMA methods under continuous baseline distributions such as a normal distribution. Lucas *et al.* [60] and Yashchin [116] showed that if a shift size is the same as a standard deviation, the CUSUM performs better than the EWMA. Srivastava and Wu [96] suggested that the EWMA is less efficient than the CUSUM under stationary conditions. Other studies showed that in a two-sided test, the performance of the EWMA is as good as that of the CUSUM (Srivastava and Wu [97]). Recently, Joner *et al.* [46] compared scan statistic methods with CUSUM for Bernoulli observations and concluded that based on a steady-state ARL, Bernoulli-based CUSUM charts performed better than Bernoulli scan statistic methods.

2.3.4 The Scan Statistics

The scan statistic procedure is defined by $T^S = \inf\{n \geq 1 | S_n \geq h\}$, where

$$S_n = \max_{1 \leq i \leq n} \{ \sum_{j=i-m+1}^{i} Y_j \}, \tag{7}$$

where m is the scanning window size, and n is the current time point. We assume that $Y_0 = Y_{-1} = \ldots = Y_{-m} = 0$. The scan statistic is mathematically related to unweighted moving average $M_n = \sum_{j=n-m+1}^{n} Y_j$ in the sequential change point detection problem (Joner $et\ al.$ [46]). The unweighted moving average for monitoring a mean is the sample mean of the observations in the most recent samples.

The scan statistic has been conventionally used for the detection of a change in the off-line setting (Glaz $et\ al.$ [29], Glaz $et\ al.$ [30], and Balakrishman $et\ al.$ [5]). For the sequential change point detection problem, the scan statistic can be modified for the detection of a change (Joner *et al.* [46]). A comprehensive review of the use of the scan statistic

for the sequential change point detection problem can be found in Woodall *et al.* [111]. Ismail *et al.* [41] and Naus *et al.* [68] proposed scan statistic methods for monitoring in the Bernoulli-temporal and exponential-temporal cases. Joner $et\ al.$ [46] investigated the performance of the scan statistic with the cumulative sum (CUSUM) chart under a Bernoulli distribution, and they demonstrated that the Bernoulli CUSUM of Reynolds *et al.* [80] performs better than the scan statistics of Naus *et al.* [68] when the rate is a sustained step function. Naus *et al.* [68] applied the scan statistics to the prospective monitoring in which the background rate is not constant. For the non-constant background case, Grigg *et al.* [31] provides an excellent review of risk-adjusted monitoring.

2.4 Our Data Sets

2.4.1 Male Thyroid Cancer in New Mexico

The specific data used in Chapters 3, 4, and 5 came from a data set of male thyroid cancer cases (with malignant behavior) in New Mexico during 1973-2005, which has been studied before in biosurveillance literature in other contexts; see, for example, [52] and [94]. The data set is available from the Surveillance, Epidemiology, and End Results (SEER) Program at the National Cancer Institute(www.seer.cancer.gov/data). The SEER program collects information on cancer incidence, mortality, and survival data from population-based cancer registries in the United States.

Three time series data of male thyroid cancer in New Mexico during 1973-2005 are in Figure 1. The first panel of the figure plots the total number of male thyroid cancers (Y axis) over years (X axis), and the second panel illustrates the trend of male population (Y axis). The third panel illustrates the crude incidence per 100,000 population (Y axis) over years (X axis).

2.4.2 Female Breast Cancer in New Hempshire

The specific data used in Chapters 6 came from a data set of female breast cancer in New Hampshire during 1968-1994, which has been studied before by Rogerson and Yamada

[88]. We obtained the annual data set of the cancer from the Compressed Mortality File from the CDC. In the chapter 6, we studied the mortality of female breast cancer in New Hampshire. Figure 2 plots the motality per 10,000 for female breast cancer in New Hampshire.

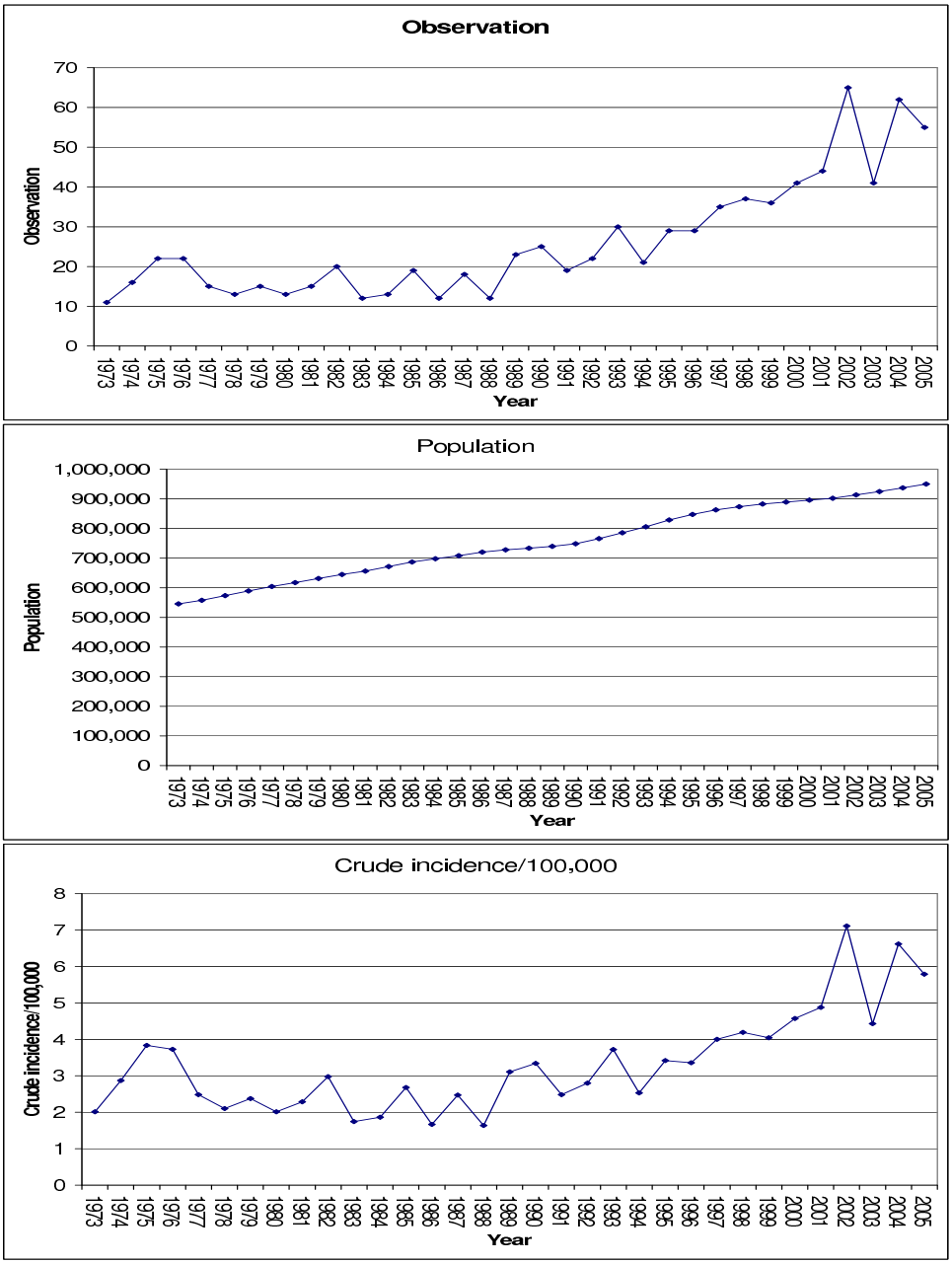


Figure 1: The three time series data related to male thyroid cancer in New Mexico during 1973-2005 illustrate the following: the total number of male thyroid cancers over years (the top panel), the trend of male population (the middle panel), and the crude incidence per 100,000 population (the bottom panel).

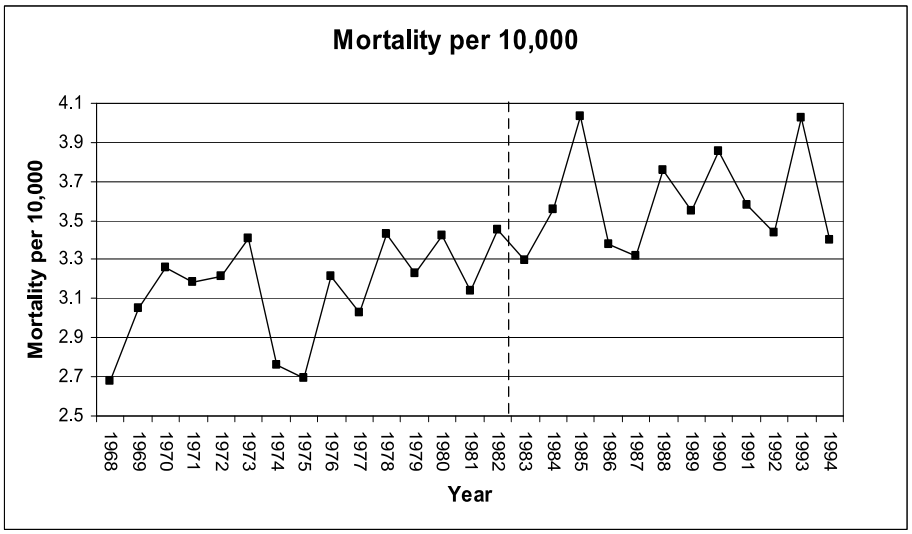


Figure 2: Mortality per 10,000 for female breast cancer in New Hampshire

CHAPTER III

A COMPARISON OF CUSUM, EWMA, AND SCAN STATISTICS FOR THE DETECTION OF RATE INCREASES IN HOMOGENEOUS POISSON DATA

In this chapter, we discuss the detection of a rate increase in Poisson count data. This chapter contains only excerpts from Han, Tsui, Ariyajunyab, and Kim [32], which was published in *Quality and Reliability Engineering International*, 2010, Vol. 26, Page 279–289. The excerpt format is based on the Copyright Transfer Agreement (2009) in Wiley Blackwell, and the excerpts have undergone several modifications in grammar.

3.1 Introduction

The objective of health surveillance, especially syndromic surveillance, is to detect a change in the incidence of natural outbreaks or bioterrorism and to issue an emergency alarm as soon as possible (Hutwagner *et al.* [40], Fricker *et al.* [25], and Fricker *et al.* [24]). Detection of such changes is based on count data, such as a count of respiratory diagnoses from civilian office visits, or measurement data such as emergency department (ED) visits, sales of over-the-counter (OTC) remedies, and the number of visits to military clinics (Rolka *et al.* [89]). These count data, observed sequentially, are often assumed to follow a certain discrete distribution such as the Poisson distribution. If an undesirable event occurs, the mean of the Poisson distribution will change.

The popular methods for detecting any such change include the scan statistic methods in biosurveillance, and the cumulative sum (CUSUM) and exponential weighted moving average (EWMA) charts in engineering statistical process control. Moustakides [67] proved that the CUSUM method is exactly optimal when Lorden's performance measures are used.

However, Lorden's criteria rate the performance of detection methods in extreme situations that rarely occur. More realistically, general and reasonable performance such as the conditional expected delay should be considered. This chapter compares the conditional expected delay among the scan statistic, CUSUM, and EWMA methods by evaluating their performance in detecting increases in rates under the discrete Poisson distribution.

The outline of this chapter is as follows: Section 3.2 describes the problem. Section 3.3 briefly reviews the Poisson version of the scan statistic, CUSUM, and EWMA methods, and Section 3.4 compares the performance of these three detection methods under various simulation scenarios. Section 3.5 presents a real application study for the detection of male thyroid cancer data in New Mexico in the United States. Section 3.6 presents the concluding remarks.

3.2 Problem Formulation

In Poisson count data, we are interested in monitoring and detecting a shift in the rate of occurrences. Assume that we observe a sequence of independent Poisson random variables over time, $\{Y_1, Y_2, \ldots\}$, in which the true mean of Y_i is μ_i. Under the normal state, $\mu_i = \lambda_0$. After an undesired event occurs at an *unknown* time ν, the values of the μ_i's change from λ_0 to λ_1. In other words, for some $\nu \geq 1$, $Y_1, \ldots, Y_{\nu-1}$ are Poisson random variables with a mean of λ_0 whereas $Y_\nu, Y_{\nu+1}, \ldots$ are Poisson random variables with a mean of λ_1. The main goal here is to detect an increase in the rate (λ) as soon as possible after an undesirable event occurs. This problem can be formulated based on the following hypothesis testing problem:

$$H: \ \mu_i = \lambda_0 \quad \text{for all } i \geq 1 \text{ (i.e., no change)}$$

against the composite alternative hypothesis

$$K: \ \mu_i = \begin{cases} \lambda_0, & \text{if } 1 \leq i \leq \nu - 1 \\ \lambda_1, & \text{if } i \geq \nu \end{cases} \quad \text{for some } \textit{unknown } \nu \geq 1 \text{ (i.e., a change).}$$

This hypothesis test (H versus K) is conducted each time based on a sequence of independent Poisson random variables. The null hypothesis H states that there is no change, which indicates that $\mu_i = \lambda_0$ until the most recent time point in monitoring. The alternative hypothesis K states a change occurred at an unknown time ν and the values of the μ_i's changed from λ_0 to λ_1 at that time.

For comparison of the detection methods in the present study, we primarily use $CED(\nu, \lambda_1)$ for each method. That is, CED's are compared by graphical justification at each point in time that a change occurs ν. Further, we considered different points of time for change ν in $CED^u(\nu, \lambda_1)$. We focused on comparing three widely used detection methods based on the scan statistic, CUSUM, and EWMA methods, by investigating their behavior in terms of $CED^u(\nu, \lambda_1)$ for different ν's and a certain range of shift sizes λ_1 subject to

$$ARL_0^u \geq \gamma, \tag{8}$$

in which γ is the lower boundary of the target.

3.3 Detection Methods

3.3.1 Scan Statistics

A scan statistic generates an alarm at time T^S, which is the first time $n(\geq 1)$ such that the scan statistic (S_n) exceeds threshold $k(\geq 1)$. Consequently, S_n is defined by

$$S_n \;=\; \max_{1 \leq i \leq n} \{ \sum_{j=i-m+1}^{i} Y_j \}, \tag{9}$$

where Y_j is the observation at time j, n is the current time point, and m is the fixed window size. k is a pre-specified threshold chosen to satisfy the constraint (8). Conventionally, we assume that $Y_0 = Y_{-1} = \ldots = Y_{-m} = 0$.

The scan statistic S_n has been used extensively in many areas of off-line decision problems (for example, Glaz $et\ al.$ [29, 30]). However, the direct use of scan statistics in (9) for online monitoring is inefficient. To efficiently implement this methodology, scan statistics

can be modified as follows:

$$S_n = \max(S_{n-1}, M_n) \quad \text{with} \tag{10}$$

$$M_n = \sum_{j=n-m+1}^{n} Y_j. \tag{11}$$

Instead of using S_n, we only need to monitor M_n. This modified scan statistic produces an alarm at time T^M, which is the first time n such that $M_n = \sum_{j=n-m+1}^{n} Y_j \geq k$.

To show that T^M is equivalent to T^S for k, first note that $T^S \leq T^M$ because the scan statistic S_n includes the statistic M_n. We can also show that $T^M \leq T^S$ as follows: If T^S issues an alarm at time n, S_{n-1} should be less than k, and S_n should be greater than or equal to k. Because $S_n = \max(S_{n-1}, M_n)$, M_n is greater than or equal to k, and T^M sets off an alarm at time n or earlier, implying that $T^M \leq T^S$. Combining the above statements yields an alarm time of $T^S = \inf\{n \geq 1 : S_n \geq k\}$, which is equivalent to $T^M = \inf\{n \geq 1 : M_n \geq k\}$, which is known as an unweighted moving average (Montgomery [64]).

3.3.2 Cumulative Sum Charts

The Poisson CUSUM chart (Brook *et al.* [12]) triggers an alarm at time T^C when the CUSUM statistic C_n exceeds h. $C_n(n \geq 1)$ can be recursively calculated by

$$C_n = \max\{0, C_{n-1} + Y_n - r\} \quad \text{with } r = \frac{\lambda_1^* - \lambda_0^*}{\ln \frac{\lambda_1^*}{\lambda_0^*}}, \tag{12}$$

where λ_0^* and λ_1^* are target in-control and out-of-control parameters, and $C_0 = 0$.

The CUSUM can be considered as a scan statistic method with a variable window size for the following reasons. First, the CUSUM chart can be interpreted from an off-line hypothesis testing viewpoint as generalized likelihood ratios in terms of changes in points of time (Basseville *et al.* [7]). More precisely, in detecting a rate change (from λ_0 to λ_1) from the distribution f_λ for the first n observations $Y_1, \ldots, Y_n$, we encounter the problem of testing H :*no change occurs* versus K : *a change occurred and* $1 \leq \nu \leq n$. Based on

this, the statistic W_n can be obtained from the following maximum of the likelihood ratio with variable time windows:

$$W_n = \max_{1 \le \nu \le n} \{ \sum_{j=\nu}^{n} \ln \frac{f_{\lambda_1}(Y_j)}{f_{\lambda_0}(Y_j)} \}. \tag{13}$$

Next, the maximum of the likelihood ratio in (13) is equivalent to the scan statistic method derived from the likelihood ratio within variable time windows. We may define scan statistics with variable window size as

$$V_n = \max_{1 \le \nu \le i \le n} \{ \sum_{j=\nu}^{i} \ln \frac{f_{\lambda_1}(Y_j)}{f_{\lambda_0}(Y_j)} \}, \tag{14}$$

which becomes scan statistics with a fixed window size if $\nu = i - m + 1$. It is easy to see that V_n satisfies

$$V_n = \max(V_{n-1}, W_n). \tag{15}$$

For $h \ge 0$, the alarm time $T^V = \inf\{n \ge 1 : V_n \ge h\}$ is equivalent to the corresponding alarm times based on W_n. Hence, the CUSUM chart can also be thought of as the scan statistic method with a variable window size for the online monitoring [94].

3.3.3 Exponential Weighted Moving Average Charts

The EWMA chart for Poisson data, proposed by Borror *et al.* [11], causes an alarm at time $T^E =$ first n such that $E_n \ge b$. The EWMA statistic E_n can be recursively calculated by

$$E_n = \alpha Y_n + (1 - \alpha) E_{n-1}, \tag{16}$$

where $0 < \alpha \le 1$, and $E_0 = E[Y]$.

For the early detection of a change, Montgomery [64] suggests an exact control limit, but we used a constant threshold in this study because the other two methods also use constant thresholds.

3.4 Simulation Study

A simulation study was conducted to explore the detection ability of the three methods (scan statistics, CUSUM, and EWMA) for Poisson count data and to compare their performance under various scenarios. Our simulation is motivated by male thyroid cancer data in New Mexico (Kulldorff [52]).

In our simulation, we set ARL_0 as close to 1,500 as possible without going below it. In other words, for all detection methods, the corresponding parameters were chosen so that the methods generate a false alarm not less than once every 1,500 time periods under the baseline incidence rate.

The rate $\lambda_0 = 1.4$ is the same baseline rate used for the spatiotemporal application in Sonesson [94]. Next, we decided to target shift sizes λ_1^*=1.75, 2.1, 2.45, and 2.8 indicating 25%, 50%, 75%, and 100% larger than λ_0, respectively. We searched the parameters and thresholds for the targeted ARL_0 based on 1,600,000 replications. Further, we simulated $CED(\nu, \lambda_1)$ for different changes in points of time and shift sizes based on 50,000 replicates.

3.4.1 Parameter Selection for Target ARL_0

To test the scan statistic methods, we obtained a set of parameters (m, k) that yield ARL_0 as close to 1,500 as possible without going below it. In the in-control state, if m decreased, ARL_0 increased because H_0 is hard to reject. The performance of the scan statistic methods depends on both m and k. Different m and k should be chosen for different shift sizes λ_1. Here is how we chose m and k for various specified shift sizes. We first tried to find all possible combinations of m and k that result in the target ARL_0. Next, for each specified shift size (λ_1), we selected combinations of m and k values that minimize ARL_1. Table 1 displays a set of the parameters for changes in target size λ_1^* and the corresponding ARL_0. The values in parentheses next to ARL_0 represent standard errors of run length values. Note that the values of m in Table 1 are the optimal choice of m in terms of ARL_1 for a

given shift size λ_1.

For the CUSUM charts, if λ_1^* in the CUSUM formula in (12) is equal to λ_1, then it yields a small ARL_0. First, we found the smallest threshold h under $\lambda_1^* = \lambda_1$ for the target ARL_0. However, the CUSUM statistic is discrete for Poisson observations. As a consequence, ARL_0 is a step function with respect to h (appendix). Table 2 shows a set of the CUSUM parameters for true shift sizes λ_1 ($= \lambda_1^*$), h, and the corresponding ARL_0.

The parameters for the EWMA charts are the weighting coefficient (α) and the threshold (b). The performance of the EWMA charts depends on both α and b. The way to select the parameters α and b is similar to the procedure used in the scan statistic methods. We first tried to find all the possible combinations of α and b that result in ARL_0's that are close to the ARL_0's in the CUSUM. Next, for each specified shift size (λ_1), we selected a set of α and b values that minimize ARL_1. Table 3 shows a set of EWMA parameters for change in target size λ_1^* and the corresponding ARL_0's.

3.4.2 Comparison of $CED(\nu, \lambda_1)$s at Different Points of Time for Changes ν at Fixed Shift Size λ_1

We first compared the three detection methods by considering different points of time at which change occurred ν. To present the simulation results efficiently, we considered the following finite number of time points: ν= 1, 2, 3, ... , 48, 49, 50. Note that most of $CED(\nu, \lambda_1)$s converge as early as 50. Figure 3 shows the resulting CED values of the scan statistics, CUSUM, and EWMA for changes at different points in time (ν), given four different shift sizes (λ_1). It can be seen that the values of $CED(\nu, \lambda_1)$ for the CUSUM and EWMA are consistently smaller than those for the scan statistics, demonstrating that the CUSUM and EWMA are more efficient than the scan statistics at detecting the points at which changes occur.

Furthermore, we generally observed that the CUSUM performed slightly better than the EWMA for a large shift in size and changes at later times. However, the EWMA performed slightly better than the CUSUM for smaller shifts and changes at early points in time.

Each detection method produced the maximum CED when $\nu = 1$. In scan statistic methods, the $CED(\nu, \lambda_1)$ tends to rapidly decrease when a change occurs at early time points and converges after $\nu \geq m$ because the scan statistics use a fixed window size m. The values of CED in the CUSUM and EWMA do not significantly change over the time change points, implying that they are more robust than the scan statistics with respect to the time change points. In addition, the EWMA is more robust than the CUSUM with respect to the point time at which a change occurs.

3.4.3 Comparison of $CED(\nu, \lambda_1)$s Under Different Shift Sizes

In the previous section, we investigated the performance of the three detection methods when λ_1 is known. This section contains the results of our investigation of their performance when λ_1 is unknown. In practice, if λ_1 is unknown, one might choose a set of parameters to detect the targeted shift size λ_1^* and investigate the pattern of CEDs under a different true shift size λ_1 at a fixed change in time $\nu = \nu^*$, where ν^* is some constant time point.

A comparison of the CEDs of the different methods may reveal a scaling issue. In order to address this scaling problem, we used a scaled version of the CED value that can be computed by dividing the $CED^u(\nu = \nu^*, \lambda_1)$ in each detection method (u) by the $CED(\nu = \nu^*, \lambda_1)$ of the optimal CUSUM (i.e., $\frac{CED^u(\nu=\nu^*,\lambda_1)}{CED^{optimal.CUSUM}(\nu=\nu^*,\lambda_1)}$).

Figures 4 and 5 display the scaled CED values of the scan statistic, CUSUM, and EWMA methods over the different values of λ_1 given $\nu^* = 1$ and $\nu^* = 1500$, respectively. Both figures showed that the CUSUM and EWMA uniformly produced smaller scaled CED values than the scan statistics, implying that the CUSUM and EWMA are superior to the scan statistics for the rapid detection of shifts of various sizes at both early (i.e., $\nu = 1$) and later (i.e., $\nu = 1, 500$) points in time when a change occurs.

In comparisons of the EWMA and CUSUM, Figures 4 and 5 indicated that the EWMA tended to perform better than the CUSUM for a small shift with an early change in time

while the CUSUM tended to perform better than the EWMA for a large shift with a late change in time.

3.5 Example: The Detection of Increased Incidence in Male Thyroid Cancer

In addition to the simulation results, we considered how to apply these detection methods in order to assess their practical applications. Our dataset contains the incidence of male thyroid cancer in New Mexico in Section 2.4. Figure 6 plots the annual incidence of thyroid cancer per 100,000 men.

The main goal of this application is to detect a change in rates as early as possible. It can be seen from Figure 6 that the rate increases after 1989 or so, and therefore, we assume that no shift occur between 1973 and 1988. We used this steady-state period to estimate the baseline rate λ_0, which is $\hat{\lambda_0} \approx 2$. We tried to detect a 25% increase of λ_0 to λ_1, which is equivalent to the targeted shift size λ_1^*=2.5.

We determined the parameters of each detection method as we did in the simulation (see Section 3.4.1). The target ARL_0 was set to 1,000. The parameters were λ_0^*=2 and λ_1^*=2.5 for the CUSUM, m=37 for scan statistics, and $\alpha = 0.02$ for the EWMA. Consequently, the thresholds of the three methods for target ARL_0 are 97 (scan statistics), 16.6 (CUSUM), and 2.33 (EWMA).

Figure 7 shows the statistics from the three detection methods over time. In order to ensure the comparability of the different methods and use the same threshold, we adjusted the values of the statistics of the scan statistic and the CUSUM methods by dividing them by 41.63 and 7.124, respectively.

It can be observed that the scan statistic, EWMA, and CUSUM methods trigger an alarm in 2004, 1999, and 2000, respectively. Assuming that the early detection is desirable, the EWMA and CUSUM triggered an alarm faster than scan statistics, which is consistent with our simulation results. Nevertheless, we cannot make a concrete conclusion about which method performs the best because this real data set lacks information about when the

shift actually occurred. The intention of this case study is to show that the three detection methods discussed here produce different results, but the overall result is consistent with the comparative performances determined in the simulation.

3.6 Conclusions

This chapter has investigated the properties of the scan statistic, CUSUM, and EWMA methods when their observations follow the Poisson distribution and compared the performance of the three methods through simulation and a case study.

The results showed that the CUSUM and EMWA charts outperformed the scan statistic methods in Poisson cases. The simulation study revealed that the CUSUM and EWMA charts were better than the scan statistic method across all possible points of time for change and across a composite range of shift sizes λ_1. Given targeted shift sizes $\lambda_1^*(=\lambda_1)$, the CUSUM and EWMA charts outperformed the scan statistic methods across all the points of time at which a change occurred. In addition, given the best methods for a targeted shift size λ_1^*, the CUSUM and EWMA charts were uniformly better than the scan statistic methods over true shift sizes λ_1.

We also compared the CUSUM charts with the EWMA charts and obtained some interesting results from the simulation. The EWMA charts were slightly better than the CUSUM charts for a small shift size with changes of earlier points in time, and the CUSUM charts were better than the EWMA charts for a large shift size with changes of later points in time.

Appendix: Obtaining Target ARL_0 from Discrete Statistics

We explain why $ARL_0 = E[T|\nu = \infty]$ in the CUSUM charts is a step function with respect to h. The condition sufficient for the statement is that if S_i is discrete and $S_{(j)}$ is the order statistics of $\{S_i, i = 1, 2, ...\}$, all $T(h)$ are identical for $h \in (S_{(j-1)}, S_{(j)})$, where $T(h) = \inf_n\{n : S_n \geq h\}$.

Suppose that $S_n \geq S_{(j)}$. Consequently, $S_n \geq h$, because $S_{(j-1)} < h < S_{(j)}$. Thus,

$$T(S_{(j)}) = \inf_n\{n : S_n \geq S_{(j)}\} \geq T(h) = \inf_n\{n : S_n \geq h\} \text{ for } h \in (S_{(j-1)}, S_{(j)}) \quad (17)$$

If $S_n \geq h$, then $S_n > S_{(j-1)}$. S_n is discrete, and $S_{(j-1)}$ and $S_{(j)}$ are adjacent order statistics. This results in $S_n \geq S_{(j)}$. Thus,

$$T(S_{(j)}) = \inf_n\{n : S_n \geq S_{(j)}\} \leq T(h) = \inf_n\{n : S_n \geq h\} \text{ for } h \in (S_{(j-1)}, S_{(j)}) \quad (18)$$

By (17) and (18), $T(h) = T(S_{(j)})$ for all $h \in (S_{(j-1)}, S_{(j)})$. Therefore, $T(h)$ is identical for all $h \in (S_{(j-1)}, S_{(j)})$.

Table 1: A set of the parameters of the scan statistic method with different shift sizes given the targeted ARL_0=1500

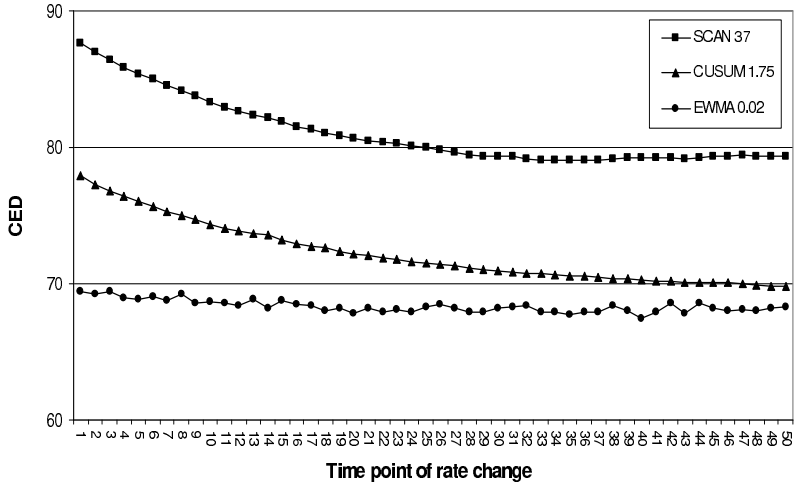
λ_1 (ν=1)	m	k	ARL_0 (s.e.)
1.75	37	72	1535.61 (1.19)
2.1	21	46	1500.57 (1.17)
2.45	14	34	1504.16 (1.18)
2.8	8	23	1519.24 (1.20)
3.15	5	17	1579.68 (1.25)

Table 2: A set of the parameters of the CUSUM charts with different shift sizes given the targeted ARL_0=1500

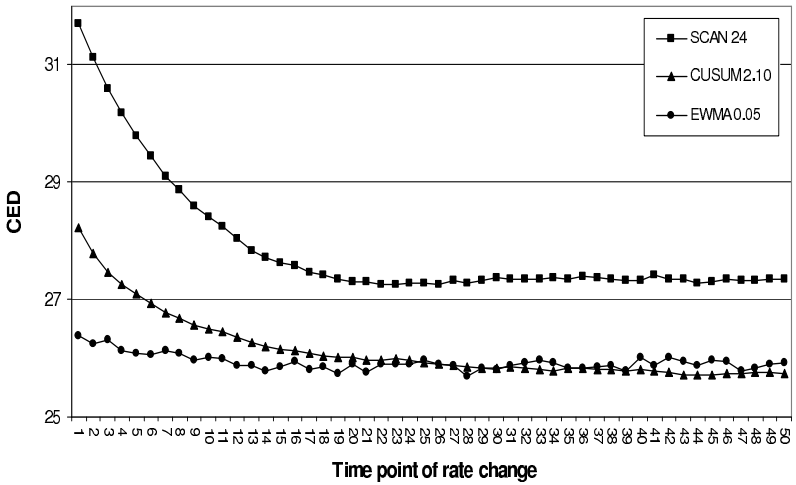
λ_1 (ν=1)	λ_1^*	h	ARL_0 (s.e.)
1.75	1.75	17.15	1547.35 (1.19)
2.10	2.10	11.68	1550.26 (1.21)
2.45	2.45	9.12	1533.51 (1.20)
2.80	2.80	7.8419	1576.68 (1.24)
3.15	3.15	6.7	1538.29 (1.21)

Table 3: A set of the parameters in the EWMA charts with different shift sizes given the targeted ARL_0=1500

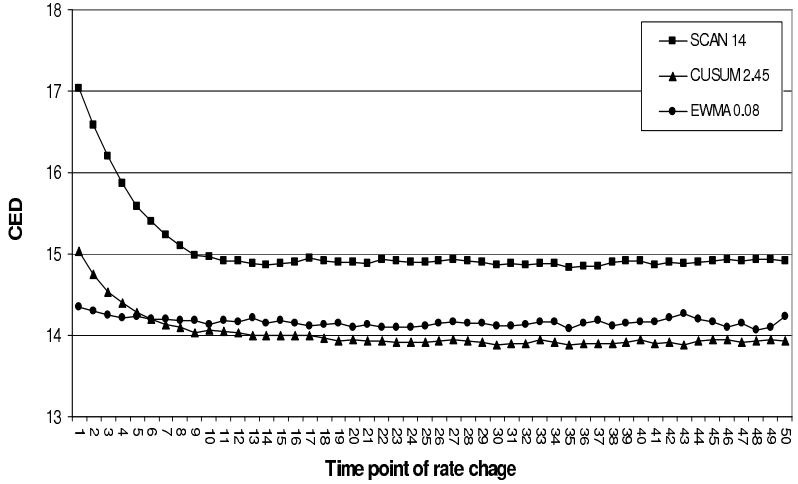
λ_1 (ν=1)	α	b	ARL_0 (s.e.)
1.75	0.02	1.7038	1547.81 (1.20)
2.10	0.05	1.9651	1551.12 (1.22)
2.45	0.08	2.1720	1533.37 (1.20)
2.80	0.12	2.4155	1576.19 (1.24)
3.15	0.17	2.6792	1538.09 (1.21)



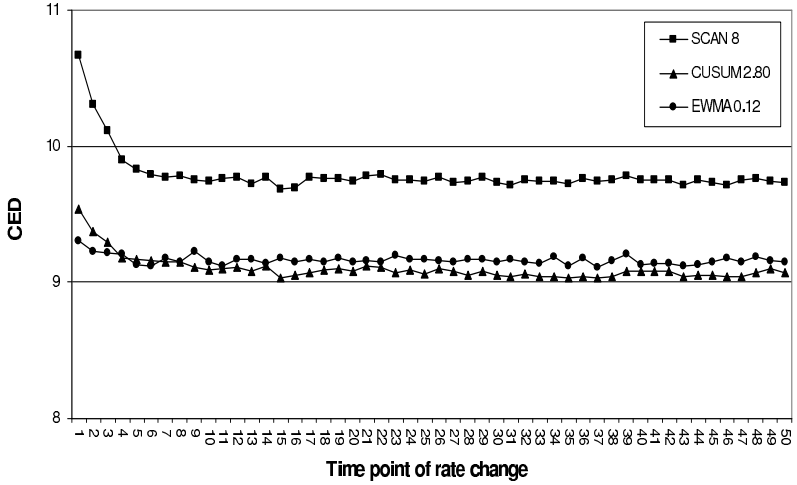
The target and true shift size $\lambda_1^*=\lambda_1=1.75$



The target and true shift size $\lambda_1^*=\lambda_1=2.1$

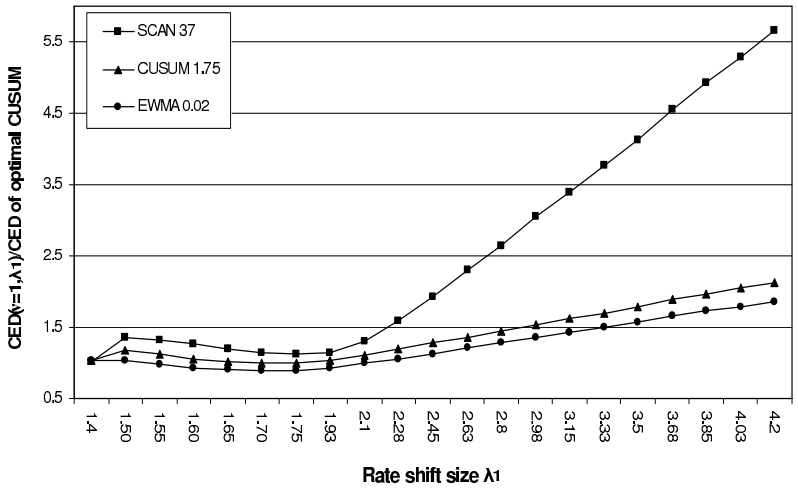


The target and true shift size $\lambda_1^*=\lambda_1=2.45$

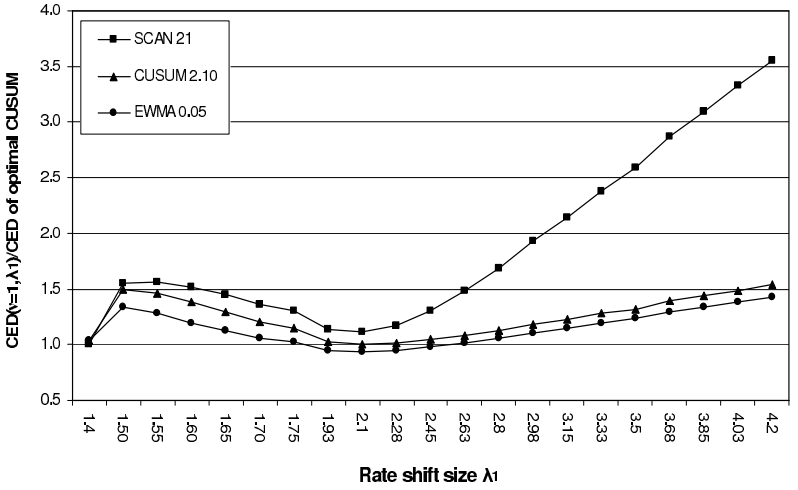


The target and true shift size $\lambda_1^*=\lambda_1=2.8$

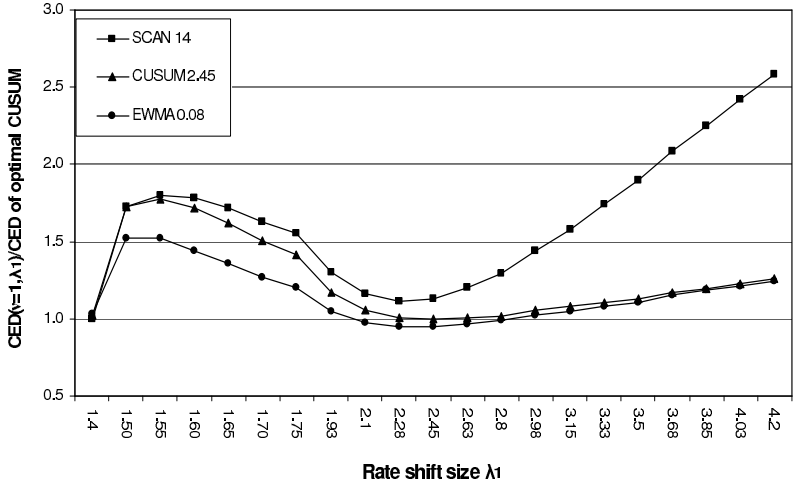
Figure 3: Comparison of the three detection methods with optimal parameters for given shift size λ_1 across different points of time for change ν. The x-axis represents time point of rate change (νs) and the y-axis represents $CED^u(\nu, \lambda_1)$s.



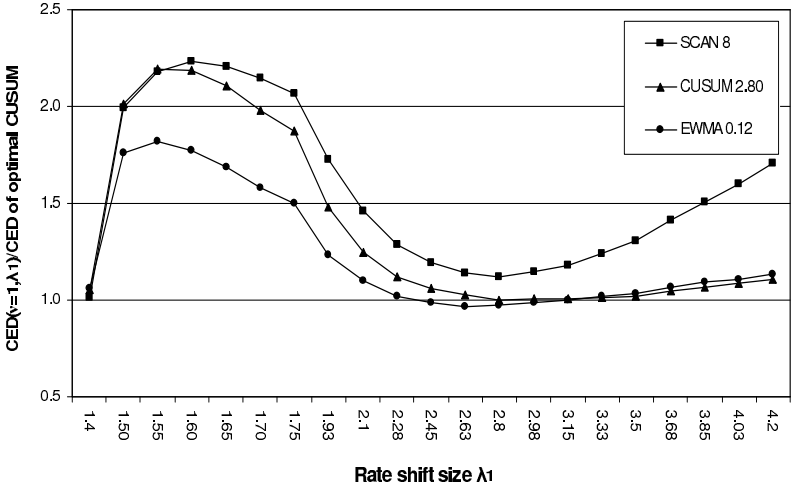
The target shift size λ_1^*=1.75



The target shift size λ_1^*=2.1

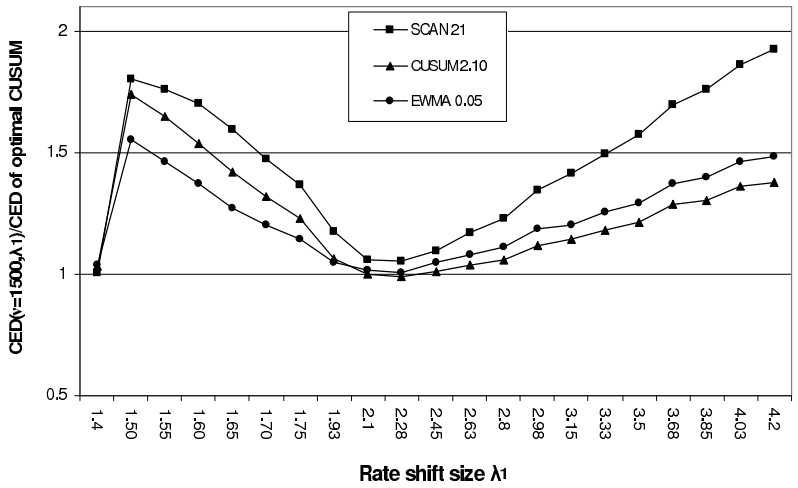


The target shift size λ_1^*=2.45

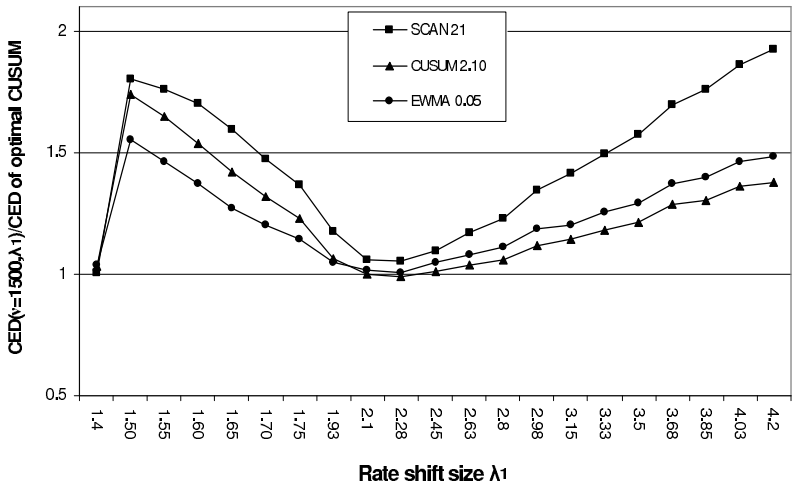


The target shift size λ_1^*=2.8

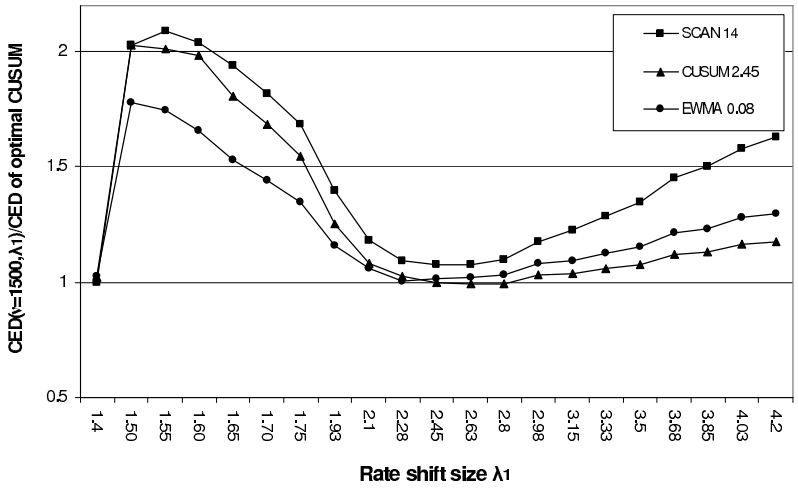
Figure 4: Comparison of the three detection methods with optimal parameters for a target shift size λ_1^* (ν=1). Sets of parameters for the scan statistics, CUSUM, and EWMA for a target size λ_1^* and change times ν=1 were used. The x-axis represents a true shift size (λ_1) and the y-axis represents $\frac{CED(\nu=1,\lambda_1)}{CED^{optimal.CUSUM}(\nu=1,\lambda_1)}$.



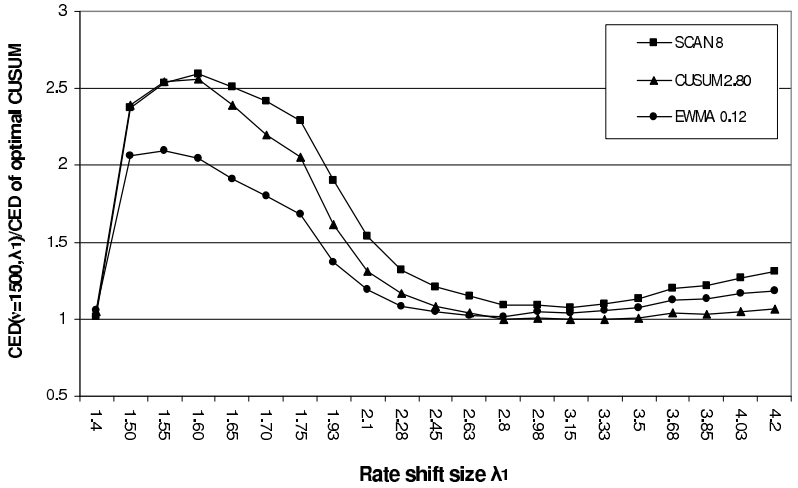
The target shift size λ_1^*=1.75



The target shift size λ_1^*=2.1



The target shift size λ_1^*=2.45



The target shift size λ_1^*=2.8

Figure 5: Comparison of the three detection methods with optimal parameters for a target shift size λ_1^* (ν=1,500). Sets of parameters for the scan statistics, CUSUM, and EWMA for a target size λ_1^* and change times ν=1,500 were used. The x-axis represents a true shift size (λ_1) and the y-axis corresponds to $\frac{CED(\nu=1{,}500,\lambda_1)}{CED^{optimal.CUSUM}(\nu=1{,}500,\lambda_1)}$.

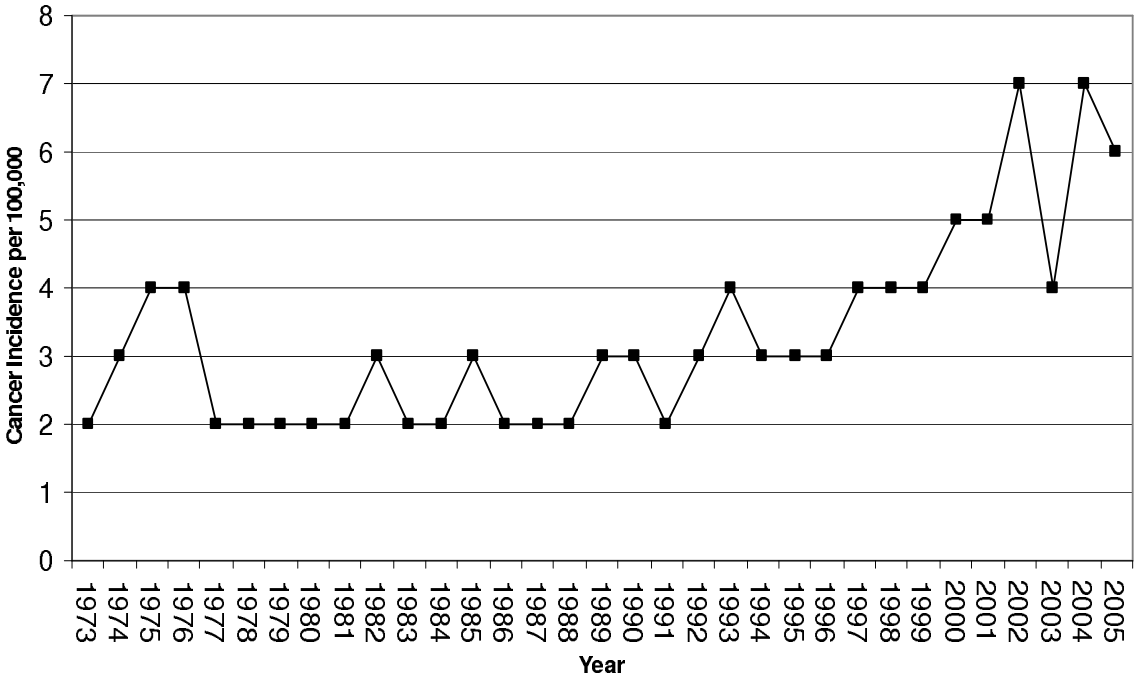


Figure 6: The trend of male thyroid cancer incidence between 1973 and 2005.

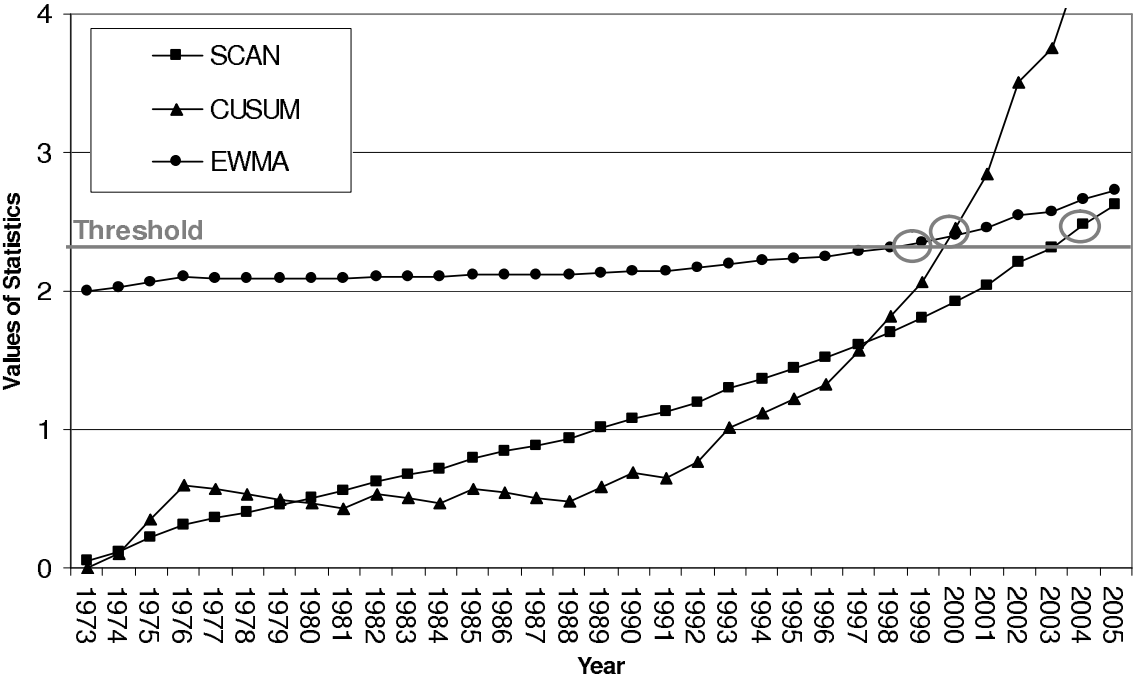


Figure 7: Plots of statistics of the scan statistic method with m=37, the CUSUM chart with λ_0^*=2 and λ_1^*=2.5, the EWMA chart with α=0.02. The circle indicates the first time point when each method triggers an alarm.

CHAPTER IV

EARLY DETECTION OF RATE INCREASES IN NON-HOMOGENEOUS POISSON DATA

In the problem of health care and bio surveillance, the sample size (often, the population size) is usually not homogeneous. For such a problem, the main goal is to detect the rate increase at an unknown time point by taking into account the population size. Several temporal surveillance methods for non-homogeneous sample sizes have been suggested and studied. Ryan *et al.* [90] reviewed the existing detection methods for non-homogeneous populations based on likelihood ratios. Montgomery [64] mentioned that the Shewhart u-chart is an attribute chart that handles non-homogeneous sample sizes. Several researchers modified the CUSUM to deal with the non-homogeneous sample size. Yashchin [115] and Hawkins *et al.* [35] developed the weighted CUSUM derived from likelihood ratios with variable sample sizes, in which the Poisson mean is proportional to the sample size. Sparks *et al.* [95] also suggested the weighted CUSUM, but they studied the method under a more general case of the Poisson mean.

In this chapter, we discuss the sequential change point detection problem with the non-homogeneous Poisson data and propose several detection schemes based on likelihood ratios. This chapter is the extended version of Mei, Han, and Tsui [63], which was accepted by *Statistica Sinica* for publication.

4.1 Introduction

Recall the male thyroid cancer data set in Figure 1 in Section 2.4.1. From the viewpoint of sequential change-point detection, classical theories and methods are applicable to this data set if one wants to investigate whether the yearly total number of male thyroid cancer

cases increases over time or not because such a problem can be formulated as detecting a change in the mean of a Poisson distribution.

From the biosurveillance viewpoint, however, a more interesting goal of this data set is to determine analytically whether or not the *risk* for male thyroid cancer increases over time. The term *risk* in this context essentially refers to the probability that one will develop thyroid cancer in a given year, which can be characterized by the incidence per 100,000 (male) population**;** see the plot in the bottom panel of Figure 1. However, for the problem of detecting a change in risk, the classical change-point detection theory and methods need to be adapted to take into account the effect of non-homogeneous population sizes in practical application.

The remainder of this chapter is as follows. Section 4.2 states the mathematical formulation of the problem, and Section 4.3 develops the generalized likelihood ratio (GLR)-based scheme and establishes its asymptotic optimality properties under a classical asymptotic setting. Section 4.4 proposes two families of alternative schemes that take into account the effect of population sizes. The GLR scheme and two proposed alternative schemes are applied to the male thyroid cancer data in Section 4.5, and more simulation results are reported in Section 4.6. To gain a deeper insight and to better reflect finite-sample numerical simulation results, Section 4.7 presents an asymptotic optimality theory under a new asymptotic setting and studies the corresponding asymptotic properties of the three proposed schemes. Section 4.8 contains some concluding remarks.

4.2 Mathematical Formulation

For the problem of detecting a change in the risk of male thyroid cancer, a very simplified probability model is the Poisson model in which one observes two-dimensional random vectors (l_n, Y_n) over time n, where Y_n has a Poisson distribution with a mean of $\mu_n = l_n\lambda_n$. Here l_n, Y_n, and λ_n can be thought of as the observed population size (in the unit of 100,000 population), the number of observed disease cases, and the (unobservable true) incidence

per 100,000 (male) population at the n-th year, respectively. Of course, in theory, it is better to model observations Y_n's by binomial distributions with the mean of $l_n\lambda_n$, but it is well known that the binomial distribution can be approximated by a Poisson distribution with the same mean, provided that the population size is large and the binomial probability parameter is small, so that the observed count is small relative to the population size. The data in our motivating example satisfy this requirement reasonably well, so the Poisson model is applicable. In addition, we also assume that observations Y_n's are independent conditional on the population sizes l_n's.

In the context of detecting a change in the risk of male thyroid cancer, it is assumed that the λ_n's, e.g., the incidence per 100,000 (male) population, change from one value λ_0 to another value λ_1 at some unknown time ν. When a change occurs, we want to detect it as soon as possible. Note that under our setting, we are interested in detecting only a change in risk λ_n's, and the population sizes l_n's can be either pre-specified constants or observable (possibly dependent) random variables whose distributions are nuisance parameters that are left unspecified.

Denote by $\mathbf{P}_\nu$ and $\mathbf{E}_\nu$ probabilities and expectations when the change in risk λ_n's occurs at time ν for $\nu = 1, 2, \ldots$, and denote the same by $\mathbf{P}_\infty$ and $\mathbf{E}_\infty$ when $\nu = \infty$, i.e., when there are no changes in the λ_n's. Mathematically, we would like to find a stopping time T that minimizes Lorden's criteria,

$$\overline{\mathbf{E}}_1(T) = \sup_{1\le\nu<\infty} \text{ess sup } \mathbf{E}_\nu\Big((T-\nu+1)^+|\mathcal{F}_{\nu-1}\Big)$$

subject to a constraint on the average run length to false alarm

$$\mathbf{E}_\infty(T) \ge \gamma, \tag{19}$$

for some given (large) constant γ.

4.3 The GLR Scheme and its Asymptotic Optimality Properties

Recall that the change-point detection problems can be thought of as testing the null hypothesis $H_0 : \nu = \infty$ (no change) against the composite alternative hypothesis $H_1 : 1 \leq \nu < \infty$ (a change occurs). The logarithm of the corresponding GLR statistic of the first n observations, $\{(l_i, Y_i)\}_{i=1}^n$, is given by

$$W_n = \max_{1 \leq \nu < \infty} \log \frac{d\mathbf{P}_\nu}{d\mathbf{P}_\infty}\Big((l_1, Y_1), \cdots, (l_n, Y_n)\Big).$$

Now given the l_i's, the Y_i's are conditionally independent with a conditional probability density function (pdf) $f_0(Y_i|l_i) = e^{-l_i\lambda_0}(l_i\lambda_0)^{Y_i}/(Y_i!)$ if $i < \nu$, but with a conditional pdf $f_1(Y_i|l_i) = e^{-l_i\lambda_1}(l_i\lambda_1)^{Y_i}/(Y_i!)$ if $i \geq \nu$. Moreover, the distribution of the l_n's are assumed to be the same under $\mathbf{P}_\infty$ or $\mathbf{P}_\nu$, and for the first n observations, $\{(l_i, Y_i)\}_{i=1}^n$, their $\mathbf{P}_\nu$-distribution is the same as their $\mathbf{P}_\infty$-distribution when $\nu > n$, due to the uniqueness of the pre-change distribution. Hence, the logarithm of the GLR statistic can be rewritten as

$$\begin{aligned} W_n &= \max_{1 \leq k \leq n+1} \sum_{i=k}^{n} \log \frac{f_1(Y_i|l_i)}{f_0(Y_i|l_i)} \\ &= \max_{1 \leq k \leq n+1} \sum_{i=k}^{n} \Big[Y_i \log \frac{\lambda_1}{\lambda_0} - l_i(\lambda_1 - \lambda_0)\Big], \end{aligned} \tag{20}$$

where $\sum_{i=n+1}^{n} = 0$ as per convention. Thus, under our setting, the GLR scheme raises an alarm at time

$$T_{GLR}(a) = \text{ first } n \geq 1 \text{ such that } W_n \geq a, \tag{21}$$

($= \infty$ if such n does not exist), where the constant a is chosen to satisfy the false alarm constraint in (19). For the purpose of online implementation, it is easy to see that W_n in (20) enjoys a recursive formula of the classical CUSUM statistics:

$$W_n = \max\Big\{0, W_{n-1} + \Big[Y_n \log \frac{\lambda_1}{\lambda_0} - l_n(\lambda_1 - \lambda_0)\Big]\Big\}.$$

It is very interesting to note that regardless of whether the population sizes l_n's are (observable) random variables or pre-specified constants, the form of GLR scheme $T_{GLR}(a)$ is

always the same. To gain a deeper understanding of the effects of population sizes from the theoretical viewpoint, from now on, we assume that population sizes l_n's are pre-specified constants, implying that observations Y_n's are independent but not necessarily identically distributed Poisson random variables.

The following theorem establishes the asymptotic lower bound subject to the constraint in (19):

Theorem **4.1** *Assume that as $n \to \infty$, population sizes l_i's satisfy*

$$\frac{1}{n}\sum_{i=k}^{k+n} l_i \to l^* \text{ uniformly for all } k > 0. \tag{22}$$

Then, for any stopping time $T(\gamma)$ satisfying the false alarm constraint in (19), we have

$$\overline{\mathbf{E}}_1(T(\gamma)) \geq (1+o(1))\frac{\log\gamma}{l^* I(\lambda_1,\lambda_0)} \qquad \text{as } \gamma \to \infty, \tag{23}$$

where

$$I(\lambda_1,\lambda_0) = \lambda_1 \log(\lambda_1/\lambda_0) - (\lambda_1 - \lambda_0). \tag{24}$$

The detailed proof, which is in Subsection 4.9.1, follows the same line of logic as those in either Lorden [57] or Lai [55]. For similar results in the context of linear regression for normal distributions, see Yao [114]. Now let us state the asymptotic optimality of the family of the GLR schemes.

Theorem **4.2** *(i) For the GLR scheme, we have $E_\infty(T_{GLR}(a)) \geq \exp(a)$ for all $a > 0$.*
(ii) Under the assumption (22) and the assumption that l_n is monotone, the generalized likelihood ratio method $T_{GLR}(a)$, satisfying the false alarm constraint in (19), satisfies that

$$\overline{\mathbf{E}}_1(T_{GLR}(a)) \leq \frac{(1+o(1))a}{l^* I(\lambda_1,\lambda_0)} \qquad \text{as } \gamma \to \infty, \tag{25}$$

where $I(\lambda_1,\lambda_0)$ is the information number defined in (24). Therefore, let $a = \log\gamma$, $T_{GLR}(a)$ attains the lower bound (23).

The detailed proof is in Subsection 4.9.2.

4.4 Two Alternative Methods

So far, it seems that we have "solved" the problem using our favorite GLR methods. However, do we really solve it from the practical viewpoint? Despite its favorable asymptotic optimality properties, the GLR scheme may not necessarily be as effective as one expects in application. To illustrate this claim, let us propose two ad-hoc alternative methods for the purpose of comparison.

Intuitively, the following two features of the GLR scheme seem to be inappropriate in the context of non-stationary population sizes: (i) the GLR statistic W_n in (20) assigns the same weight to the individual log-likelihood ratio statistic $\log \frac{f_1(Y_i|l_i)}{f_0(Y_i|l_i)}$ regardless of population sizes l_i's, but the Y_i's with larger population sizes l_i's seem to provide more information, and (ii) the GLR scheme $T_{GLR}(a)$ uses the constant threshold value a over time. These features motivate us to propose the following two alternative detection schemes that take into account the effects of population sizes.

The first alternative scheme is based on the quasi-log-likelihood ratio statistics, which normalize each term $\log \frac{f_1(Y_i|l_i)}{f_0(Y_i|l_i)}$ in (20) by their (conditional) variances, or equivalently by the population sizes l_i's (up to a constant). This viewpoint leads to a new detection statistic

$$\begin{aligned} \hat{W}_n &= \max_{1\le k\le n+1} \sum_{i=k}^{n} \frac{1}{l_i} \log \frac{f_1(Y_i|l_i)}{f_0(Y_i|l_i)} \\ &= \max_{1\le k\le n+1} \sum_{i=k}^{n} \Big[\frac{Y_i}{l_i} \log \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0) \Big]. \end{aligned} \tag{26}$$

Thus, for any given constant b, we can define the following weighted likelihood ratio (WLR) scheme

$$T_{WLR}(b) = \text{ first } n \ge 1 \text{ such that } \hat{W}_n \ge b. \tag{27}$$

Another motivation of the WLR scheme $T_{WLR}(b)$ in (27) is based on Y_n/l_n, a natural estimator of the *risk* or the disease rate per 100,000 population. To understand this, if we imagine that Y_n/l_n is Poisson distributed with mean λ_n (this is not true under our setting,

but we can still use it to construct detection schemes), then the problem becomes the classical problem of detecting a change in the Poisson mean from λ_0 to λ_1, and the corresponding CUSUM procedure is just the WLR scheme $T_{WLR}(b)$ in (27).

To take into account population size effects, the second alternative scheme we propose is to use the GLR-based statistic W_n in (20) but with adaptive thresholds. Ideally, one would like to use the optimal thresholds or boundaries, say, by some Bayesian or non-Bayesian argument, but such boundaries seem to be too complicated to be derived explicitly. For simplicity, here we will use the linear boundaries: $l_n c$ (see Section 8.1 below for a more fundamental explanation under a general setting). Specifically, the proposed adaptive threshold method (ATM) raises an alarm at time

$$T_{ATM}(c) = \text{ first } n \geq 1 \text{ such that } W_n \geq l_n c, \tag{28}$$

for some constant $c > 0$, where W_n is the GLR statistic in (20).

It is important to point out that when population sizes l_n's are equal to a constant $l > 0$, then the three proposed detection schemes, $T_{GLR}(a)$, $T_{WLR}(b)$, and $T_{ATM}(c)$ not only are equivalent (when $a = lb = lc$), but also hold the optimality properties of Page's CUSUM procedures proved in Moustakides (1986).

When population sizes l_n's vary, the conclusions are obscure. On the one hand, under the assumption of Theorem 4.2, population sizes l_n's converge to a constant value l^*, and thus, these three schemes seem to be *asymptotically* equivalent and asymptotically optimal. On the other hand, when population sizes l_n's vary, these three detection schemes are generally not equivalent, and they will likely have different finite-sample properties, which will be the focus of the remainder of this article.

4.5 Example Revisited

In this section, we revisit the male thyroid data in New Mexico. When we apply these three schemes to the real data set, it is surprising that the GLR scheme does not work as well as the two proposed alternative schemes (WLR and ATM).

Table 4: Estimated parameters for the population growth model in (29)

Parameter	ϕ_1	ϕ_2	ϕ_3	σ
Estimate	13.8065 ± 0.9552	11.8532 ± 3.7438	26.4037 ± 2.3127	0.0907

4.5.1 Model for Population Growth

In our application or simulations, we need a model that generates population sizes beyond the observed ones so that we can determine the threshold values of the detection schemes so as to satisfy the false alarm constraint (19). In the literature, such as Pinheiro and Douglas (2000, p. 274), it is common to model the growth curve by the following logistic model:

$$l_n = \psi(n) + \epsilon_n = \frac{\phi_1}{1 + \exp[-(n - \phi_2)/\phi_3]} + \epsilon_n, \tag{29}$$

where $E(\epsilon_n) = 0$ and $Var(\epsilon_n) = \sigma^2$. Here ϕ_1 indicates an asymptotic upper limit of the population size, ϕ_2 the middle point of t in the S-shaped curve, and ϕ_3 the scale adjustment of the time periods.

In our specific application of New Mexico, we fit this model to the observed population sizes by a nonlinear least-squares method (we treat year 1972 as time 0, and the population sizes are in units of 100,000). Using the statistical software R version 2.8.0, the estimated parameters of the logistic model for the population sizes are summarized in Table 4.

Figure 8 plots the actual observed population sizes and the estimated growth curve in New Mexico during 1973-2005. From the plot, it is easy to see that the two curves are very close to each other, implying that the logistic model is reasonable in our application. One may also wonder if population sizes increase linearly, and if so, why not just fit a linear model for population sizes? The answer depends on how large the false alarm constraint γ in (19) is. When γ is in the range of 30, the linear model may be reasonable since the observed 32 population sizes indeed seem to be linear. However, if γ is moderately large, say 300, it seems unrealistic to assume that population sizes increase linearly over 300 years, and the logistic model (29) may be more reasonable. In our simulations, we are more interested in a moderately large value of γ, and thus, we will use the mean curve

of the logistic model (29) with the estimated parameters in Table 4 to generate population sizes.

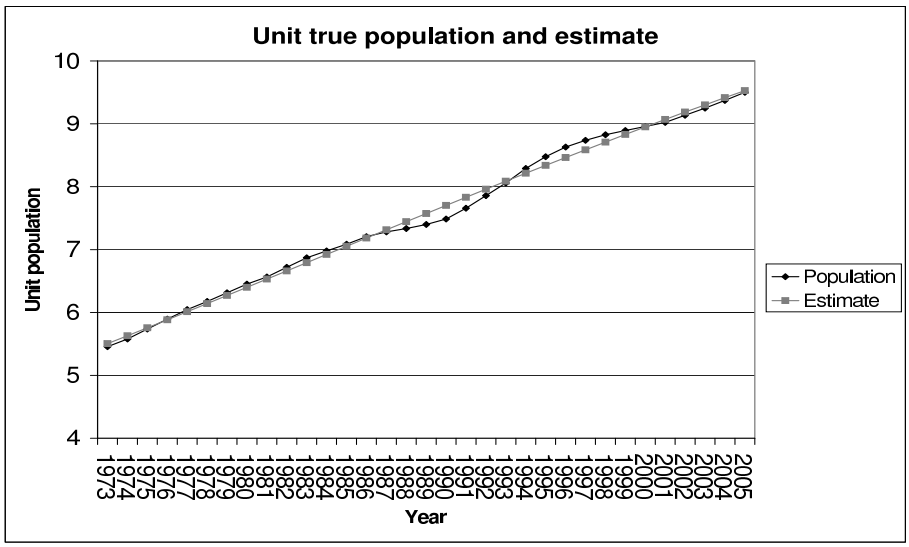


Figure 8: Population and estimate in New Mexico during 1973-2005. The plot shows the actual observed and model-estimated male population sizes in New Mexico during 1973-2005.

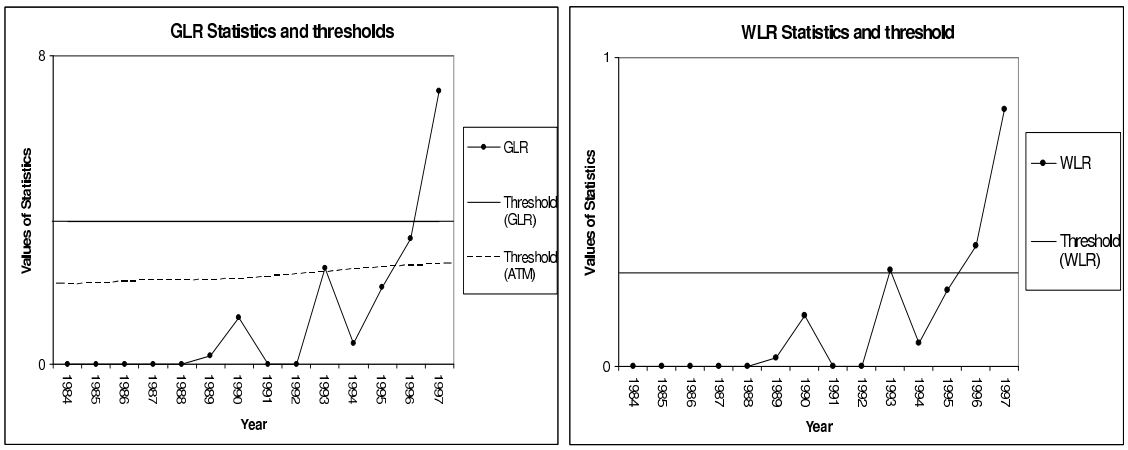


Figure 9: The *left* panel plots the GLR statistics W_n over time n, as well as the alarm boundaries of $T_{GLR}(a)$ (solid line) and $T_{ATM}(c)$ (the dotted line). The *right* panel plots the WLR statistic $\hat{W}_n$ as well as the boundary of $T_{WLR}(b)$.

4.5.2 Parameters in the Change-point Problem and Detection Schemes

In the change-point detection problem, we need to specify pre-change rate λ_0, post-change rate λ_1, and false alarm constraint γ in (19). To do so, one possible approach is to use the

time period of 1973-1983 as a training period, and then to estimate pre-change rate λ_0 and post-change rate λ_1 by the median and the maximum of the crude incidence per 100,000 during 1973-1983, respectively. For our data, we have $\lambda_0 = 2.4$ and $\lambda_1 = 3.8$.

In addition, we assume that $\gamma = 300$ in the false alarm constraint (19), i.e., on average, we want all detection schemes to raise a false alarm at least once every 300 years when the risk of cancer is $\lambda_0 = 2.4$ per 100,000 population and when no change occurs. Of course, the choices of the pre- and post-change risks, $\lambda_0 = 2.4$ and $\lambda_1 = 3.8$, and the false alarm constraint $\gamma = 300$ are intended only for the illustration purposes, and the idea can easily be extended to any other different values.

In order for the three detection schemes to satisfy the false alarm constraint (19) with $\gamma = 300$, numerical simulations show that when the population sizes are generated from the model (29) with the parameters in Table 4, the corresponding threshold values for $T_{GLR}(a)$ in (21), $T_{WLR}(b)$ in (27), and $T_{ATM}(c)$ in (28) are $a = 3.6870, b = 0.2975$, and $c = 0.2975$, respectively (based on 100,000 replicates).

4.5.3 When to Raise an Alarm

We now apply the GLR scheme and two alternative schemes, the WLR and ATM schemes, to monitor the cancer risk for the male thyroid cancer data in New Mexico, starting from the year 1984. (Recall that the time period of 1973-1983 has been used as a training period.) Figure 9 plots GLR statistic W_n in (20) and WLR statistic $\hat{W}_n$ in (26) over the time n, as well as the detection boundaries of the three proposed detection schemes when the false alarm constraint is $\gamma = 300$. From the plots, the WLR and ATM schemes, $T_{WLR}(b)$ and $T_{ATM}(c)$, trigger an alarm in the year 1993 whereas GLR scheme $T_{GLR}(a)$ raises an alarm in the year 1997. In addition, numerical simulations show that the same conclusion holds if we control the false alarm constraint at $\gamma = 100$ or 200. These results seem to suggest that for the male thyroid cancer data in New Mexico, the ad-hoc alternative schemes, $T_{WLR}(b)$ and $T_{ATM}(c)$, are better than the GLR scheme $T_{GLR}(a)$ in the sense of raising an alarm

earlier given that the rate did increase.

4.6 More Simulation Study

Of course, a single data set is not convincing to show that a statistical method is good or bad. Moreover, the truth about the real data set may be not clear-cut, as the earlier alarms can also be false alarms. In this section, we perform further simulation studies to compare the GLR scheme with two alternative schemes. In particular, we want to see whether the poor performance of the GLR scheme is a fluke or not.

In the following numerical simulation study, we will borrow values and ideas from the real cancer data set in the previous section but with two modifications. In the first, we now assume that post-change risk $\lambda_1 = 2.7$ instead of $\lambda_1 = 3.8$ in the real data set application (we still assume that the pre-change risk is $\lambda_0 = 2.4$) to investigate a smaller change rather than a larger change since a larger change may be easily detected by any reasonable methods. Indeed, in the previous section, the detection statistics W_n or $\hat{W}_n$ are 0 before crossing the detection thresholds for the WLR and ATM detection schemes, indicating that the detection delays of these methods are very small (at most 1) due to a larger change. The second modification is to control the false alarm constraint at $\gamma = 1000$ (instead of the previously used 300). We hope that a larger value of γ may lead to a larger detection delay so that we can better understand the asymptotic properties of the three detection schemes.

In our simulations below, we investigate the scenario when the population sizes are from the logistic model in (29) with different parameter values. The following three cases will be studied ($\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ and $\hat{\sigma}$ are the estimates in Table 1):

- Case A (Increasing): $l_n = \frac{\hat{\phi}_1}{1+\exp[-(n-\hat{\phi}_2)/\hat{\phi}_3]}$, which is the model we used in our real data example, and we want to investigate what happens for a smaller change (the post-change risk $\lambda_1 = 2.7$ instead of 3.8).

- Case B (Fast Increasing): $l_n = \frac{2\hat{\phi}_1}{1+\exp[-(n-\hat{\phi}_2)/(\hat{\phi}_3/3.4)]}$, where $\hat{\phi}_3/3.4$ is chosen so that population sizes l_n's will increase very fast over time to a stationary value as

compared to the real data application.

- Case C (Decreasing): $l_n = \frac{\hat{\phi_1}/2.15}{1+\exp[(n-\hat{\phi_2})/(\hat{\phi_3}/2)]} + 1$, which leads to a decrease in population sizes l_n over time n. Here, the constant 1 ensures that the population size l_n decreases to a nonzero constant.

Note that in the models of population sizes, $2\hat{\phi}_1$ in Case B and $\hat{\phi}_3/2$ in Case C are necessary to ensure that the initial population size l_0 is the same as the observed value l_0. The *top left* panel of Figure 10 plots the three different cases of population size models. In particular, Case A is the estimated curve from the real-data set, Case B corresponds to the case in which the population size quickly increases to stationary level $2\hat{\phi}_1$, and Case C is when the population size decreases rapidly to stationary value 1.

For each population size model, we first determine the detection threshold values through 100,000 replicates so that the detection schemes satisfy the false alarm constraint (19) with $\gamma \approx 1000$ (within a sampling error) and the pre-change risk $\lambda_0 = 2.4$. Then we simulate the detection delay ess sup$\mathbf{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$ at different change-point ν with the post-change risk $\lambda_1 = 2.7$. The simulated detection delays are based on 50,000 replicates.

The detection delays of the three schemes are plotted in Figure 10 for the three different cases of population size models. From Figure 10, it is interesting to note that if the population sizes are decreasing (Case C), the GLR scheme $T_{GLR}(a)$ seems to be the best (in the sense of the smallest detection delay at each change-point ν), the WLR scheme $T_{WLR}(b)$ seems to be the worst, and the ATM scheme $T_{ATM}(c)$ seems to be in-between. However, if the population sizes are increasing (Cases A and B), the order is reversed: $T_{WLR}(b)$ seems to be the best, $T_{GLR}(a)$ is the worst, and $T_{ATM}(c)$ is still in-between.

Moreover, under Lorden's worst-case detection delay $\overline{\mathbf{E}}_1(T)$ criterion, if the population sizes are increasing, $T_{GLR}(a)$ is the worst scheme whereas the performances of $T_{WLR}(b)$ and $T_{ATM}(c)$ are better. On the other hand, if the population sizes are decreasing, $T_{WLR}(b)$ is the worst scheme in the sense of the largest worst-case detection delay $\overline{\mathbf{E}}_1(T)$, whereas the performances of $T_{GLR}(a)$ and $T_{ATM}(c)$ are better. This suggests that the adaptive

threshold scheme $T_{ATM}(c)$ seems to be *robust* in the sense of small detection delays $\overline{\mathbf{E}}_1(T)$ under Lorden's criterion, regardless of whether the population sizes are increasing or decreasing.

It is also interesting to see from Figure 10 that for $T_{GLR}(a)$, the detection delays ess sup$\mathbf{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$ seem to be decreasing (increasing) as a function of the change-point ν when the populations sizes are increasing (decreasing). However, the detection delays of $T_{WLR}(b)$ seem to be an increasing (decreasing) function of the change-point ν when the populations sizes are increasing (decreasing). The pattern of its detection delay of $T_{ATM}(c)$ is unclear since it may not necessarily be a monotone function of ν. In all simulations, the three detection schemes have similar detection delays, ess sup$\mathbf{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$ when the change-point ν is large, but they have very different detection delays when the change-point ν occurs at an earlier stage.

In summary, our numerical simulations suggest that in real-world applications with prior information that population sizes are increasing or decreasing, one should use the best scheme among these three schemes, i.e., the WLR scheme for increasing population sizes and the GLR scheme for decreasing population sizes. When there is uncertainty about the trends of population sizes, one may want to use the adaptive threshold scheme $T_{ATM}(c)$ to take advantage of its robustness properties. In particular, despite its asymptotic optimality properties, the GLR scheme indeed can perform very poorly in finite-sample numerical simulations.

4.7 New Asymptotic Analysis

The main purpose of this section is to develop some theory to "explain" our numerical simulation results in the previous section. Evidently, the setting of such a theory cannot be the classical asymptotic setting in Theorem 4.2. To find an appropriate setting, let us go back to our numerical simulations. Since the GLR scheme is at times efficient and at times inefficient in our simulations, a natural reaction is to check whether the condition

(22) required in Theorem 4.2 holds or not.

On the one hand, the condition (22) holds from the purely mathematical viewpoint, since in our simulations, the population sizes are from the logistic model, in which the population sizes monotonically increase or decease to stationary value l^*. Thus, the GLR scheme should have been asymptotically optimal under the setting of our simulation study, regardless of whether the population sizes are increasing or decreasing.

On the other hand, from a computational viewpoint, our simulations violate the *spirit* of the condition (22). To be more specific, the false alarm constraint γ is only moderately large in our simulation in view of the uniform convergence in the assumption (22). That is, since the false alarm constraint γ is not too large, the post-change sample size n is generally not too large, and thus, the value of $\frac{1}{n}\sum_{i=k+1}^{k+n} l_i$ for small values of k can be very different from the corresponding value for a large value of k.

To gain a deeper understanding, let us now assume that population sizes l_n's reach stationary value l^* at some finite time ω. Then, Theorem 4.2 deals with the scenario when ω is much smaller than γ, e.g., fix ω and let the false alarm constraint γ go to ∞. However, our numerical simulations deal with a different setting in the sense that ω is comparable to the false alarm constraint γ, or they are on the same order. For instance, when fitting the logistic model to the real data, the population sizes are close to the stationary value numerically around the time $\omega \approx 120$ while the false alarm constraint $\gamma = 1000$.

Therefore, in order to reflect finite-sample numerical results, in this section, we consider a new asymptotic setting in which the population sizes reach stationary value l^* at some finite time ω, where $\omega = \omega_\gamma \leq C\gamma$ for some constant $0 < C < 1$, as the false alarm constraint γ goes to ∞. Note that under our new asymptotic setting, as $n \to \infty$, $\frac{1}{n}\sum_{i=k+1}^{k+n} l_i$ converges to l^* point-wise for each k, but this convergence is not uniform over k as required in the assumption (22) of Theorem 4.2.

To better present the results under our new asymptotic setting, the remainder of this section is divided into two subsections. Section 4.7.1 presents an asymptotic theory under

the new setting, and Section 4.7.2 reports further simulations to illustrate our theory.

4.7.1 Asymptotic Analysis

As in the classical asymptotic optimality, we will first construct an asymptotic lower bound on the detection delays as the false alarm constraint γ in (19) goes to ∞. Then we will show whether the family of schemes will attain the lower bound asymptotically.

First, let us derive the information bound. The following theorem whose proof will be postponed to Subsection 4.9.3 shows that the information bound in Theorem 4.1 is still valid under our new condition.

Theorem **4.3** *Assume that population sizes l_n's reach the stationary value l^* at some finite time ω, where $\omega = \omega_\gamma < C\gamma$ for some constant $0 < C < 1$. Then for any stopping time $T(\gamma)$ satisfying the false alarm constraint in (19), the relation (23) holds.*

Second, let us state a general result on the detection delays of the WLR schemes. The detailed proof is in Subsection 4.9.4. The result is useful for the asymptotic optimality analysis of the comparison of the three schemes.

Theorem **4.4** *For the WLR scheme, if $\inf_{n\geq 1} l_n = l_* > 0$, then as $b \to \infty$,*

$$\overline{\mathbf{E}}_1(T_{WLR}(b)) \leq \frac{b}{I(\lambda_1, \lambda_0)} + M,$$

where $I(\lambda_1, \lambda_0)$ is defined in (24) and the constant

$$M = \sqrt{1 + \frac{\lambda_1}{l_*}\Big(\lambda_1 - \frac{\lambda_1 - \lambda_0}{\log \lambda_1 - \log \lambda_0}\Big)^{-2}}.$$

Now given Theorems 4.3 and 4.4 in order to derive the asymptotic optimality properties of the GLR or WLR schemes subject to the false alarm constraint γ in (19), we need to derive the detection delays of the GLR scheme $T_{GLR}(a)$ and/or to establish the relationship between the threshold value b in the WLR scheme $T_{WLR}(b)$ and the false alarm constraint γ. Unfortunately, both are very challenging under our new asymptotic setting. In addition, it is

non-trivial to investigate the false alarm or detection delay properties of the ATM scheme. Also, recall that the uniform convergence assumption (22) is needed to derive the detection delay properties of the GLR scheme $T_{GLR}(a)$, which is intractable otherwise. Therefore, below we make two further simplifications in order to be tractable in theory under our new asymptotic setting.

The first one is to focus on two kinds of changes (on the disease risk): one is when the change occurs at time $\nu = 1$, and the other is when the change occurs at time $\nu = \omega$. The change-point $\nu = 1$ is used to indicate the detection delays for small values of change-point ν. The change-point $\nu = \omega$ is interesting because the detection delay ess sup $\mathbf{E}_\nu\big((T - \nu + 1)^+ | \mathcal{F}_{\nu-1}\big)$ is the same for all change-point $\nu \geq \omega$ since the observations Y_n's are i.i.d. with the same population sizes $l_n = l^{(1)}$ for $n \geq \omega$. This motivates us to consider the following detection delay:

$$D(T) = \max\left[\mathbf{E}_{\nu=1}(T),\ \text{ess sup}\ \mathbf{E}_\omega\Big((T - \omega + 1)^+ | \mathcal{F}_{\omega-1}\Big)\right],$$

which provides a lower bound on Lorden's worst-case detection delay $\overline{\mathbf{E}}_1(T)$.

The second simplification is to assume that initial population size l_n's are a constant $l^{(0)}$ for a reasonably long period so that we can use the classical results of Page's CUSUM procedures to derive the detection delay $\mathbf{E}_{\nu=1}(T)$ for the GLR and ATM schemes. Specifically, we now assume that population sizes can be modeled by the step function

$$l_n = \begin{cases} l^{(0)}, & \text{if } n < \omega \\ l^*, & \text{if } n \geq \omega \end{cases}, \tag{30}$$

where the change-times $\omega >> \log \gamma$ and $\omega < C\gamma$ for some $0 < C < 1$.

The step function is designed mainly to simplify the computation of the detection delays of the GLR or ATM schemes. However, the step function (30) for the population size model also raises a subtle but important issue. Intuitively, we do not want ω in (30) to be too large compared to the false alarm constant γ, since otherwise, the problem may essentially become the classical change-point detection problem with constant population

sizes $l^{(0)}$. Thus, to make the new asymptotic setting meaningful, we need to make sure that for the proposed three schemes, false alarms occur under population l^*. For instance, let us consider WLR scheme $T_{WLR}(b)$, and note that its average run length to false alarm is $\exp(l^{(0)}b)$ and $\exp(l^*b)$, respectively, when the population sizes are constant values of $l^{(0)}$ and l^*. Thus, we would like to have $\omega << \exp(l^{(0)}b)$, but $\exp(l^*b) \approx \gamma$. Combining them leads to $\omega << \gamma^{l^{(0)}/l^*}$, which obviously holds when $l^{(0)} > l^*$, but not necessarily true when $l^{(0)} \leq l^*$. Thus, some additional assumptions based on (30) are necessary for our new asymptotic setting. It turns out that the assumptions we need are that $\omega = o(\gamma^{(1-\eta)l^{(0)}/l^*})$ for some constant $0 < \eta < 1$. These additional assumptions are slightly more restrictive than what one may prefer, but they have an advantage in that they are able to derive the asymptotic optimality properties of the three proposed schemes.

The following theorem summarizes the asymptotic properties of the GLR, WLR, and ATM schemes under our new asymptotic setting. The detailed proof is in Subsection 4.9.5.

Theorem **4.5** *Assume that population sizes l_n's follow the step function in (30) such that $\inf_{n\geq 1} l_n = l_*$, and the false alarm constraint γ in (19) satisfies the relation $\log\gamma << \omega < C\gamma$ and $\omega = o(\gamma^{(1-\eta)l^{(0)}/l^*})$ for some $0 < \eta < 1$ as $\gamma \to \infty$. Then, subject to the false alarm constraint in (19), we have*

$$\begin{aligned} D[T_{GLR}(a)] &= (1+o(1))\frac{\log\gamma}{\min\{l^{(0)}, l^*\}I(\lambda_1,\lambda_0)}, \\ D[T_{WLR}(b)] &= (1+o(1))\frac{\log\gamma}{l^*I(\lambda_1,\lambda_0)} \\ D[T_{ATM}(c)] &= (1+o(1))\frac{\log\gamma}{l^*I(\lambda_1,\lambda_0)}, \end{aligned}$$

as $\gamma \longrightarrow \infty$, where $I(\lambda_1,\lambda_0)$ is defined in (24).

Finally, for our new asymptotic setting, we are now in the position to state the asymptotic optimality or sub-optimality of the WLR and GLR schemes under Lorden's worst-case detection delay criterion.

Corollary **4.1** *Under the assumption of Theorem 4.5, the WLR scheme $T_{WLR}(b)$ is asymptotically optimal under Lorden's worst case detection delay criterion $\overline{\mathbf{E}}_1(T)$, subject to the false alarm constraint in (19). Moreover, when $l^{(0)} < l^*$, the GLR scheme $T_{GLR}(a)$ is asymptotically suboptimal under Lorden's worst-case detection delay criterion.*

Proof. For the WLR scheme $T_{WLR}(b)$, by the proof of Theorem 4.5, the threshold $b \sim \log \gamma$ will be sufficient to satisfy the false alarm constraint γ in (19). Combining this relationship with Theorem 4.4 (ii) yields the detection delay

$$\overline{\mathbf{E}}_1(T_{WLR}(b)) \leq \frac{(1+o(1))\log\gamma}{l^* I(\lambda_1, \lambda_0)}.$$

Therefore, the WLR scheme attains the lower bound on the detection delays in Theorem 4.1, and thus, it is asymptotically optimal under Lorden's worst-case detection delay criterion $\overline{\mathbf{E}}_1(T)$. Meanwhile, if $l^{(0)} < l^*$, the sub-optimality properties of the GLR scheme $T_{GLR}(a)$ follow directly from Theorem 4.5, its comparison with the asymptotic optimal scheme $T_{WLR}(b)$, and the fact that $\overline{\mathbf{E}}_1(T) \geq D(T)$:

$$\lim_{\gamma\to\infty} \frac{\overline{\mathbf{E}}_1(T_{GLR}(a))}{\overline{\mathbf{E}}_1(T_{WLR}(b))} \geq \lim_{\gamma\to\infty} \frac{D(T_{GLR}(a))}{(1+o(1))\log\gamma/(l^* I(\lambda_1,\lambda_0))} = \frac{l^*}{\min(l^{(0)}, l^*)} \geq 1,$$

completing the proof. ■

The above theorems and corollary are consistent with our finite sample simulation results that when population sizes increase, the WLR scheme is the best and the GLR scheme is the worst. On the other hand, when the population sizes decrease, i.e., when $l^{(0)} > l^*$, these three detection schemes are asymptotically equivalent in the sense that $D(T) \sim \frac{\log\gamma}{l^* I(\lambda_1,\lambda_0)}$ for any of these three detection schemes. Unfortunately, it appears that our previous simulations in Section 4.6 do not support this claim. A closer look indicates that this is partly because the asymptotic optimality theorem for the WLR scheme requires that all population sizes l_n's are bounded below by $l_* > 0$. However, the lower bound l_* in our previous numerical simulations is too small ($l_* = 1$). A larger value of l_* may be more

informative to our asymptotic theory. Hence, we will conduct a further simulation below to check the case of $l^{(0)} > l^*$.

It is crucial to highlight that $T_{WLR}(b)$ and $T_{ATM}(c)$ are equalizer rules in the sense that the detection delay is the same for $\nu = 1$ or ω. However, $T_{GLR}(a)$ is not an equalizer rule. The property of an equalizer rule is essential to establish the exact optimality of the CUSUM procedure in the simplest i.i.d. models, and our results suggest that GLR scheme $T_{GLR}(a)$ may lose such an important property in the finite-sample setting when the population sizes vary.

4.7.2 Numerical Simulations

In this subsection, we conduct a numerical study when the population sizes are modeled by the step function in (30) with $\omega = 200$. We assume that the false alarm constraint is $\gamma = 1000$, so the choice of $\omega = 200$ is consistent with the assumption of our asymptotic theorem that $\log \gamma << \omega < C\gamma$ for some $0 < C < 1$. Two cases depending on whether the population sizes are increasing or decreasing will be considered: (1) Increasing: $l^{(0)} = 6$ and $l^* = 12$; and (2) Decreasing: $l^{(0)} = 12$ and $l^* = 6$. As in Section 4.6, we assume that the pre- and post-change risks are $\lambda_0 = 2.4$ and $\lambda_1 = 2.7$, respectively.

When the population sizes are modeled by the step function in (30) with $\omega = 200$, $l^{(0)} = 6$ and $l^* = 12$, our simulations show that subject to the false alarm constraint (19) with $\gamma \approx 1000$, the threshold values for $T_{GLR}(a)$ in (21), $T_{WLR}(b)$ in (27), and $T_{ATM}(c)$ in (28) are $a = 4.540, b = 0.453$, and $c = 0.452$, respectively. Meanwhile, when the population sizes are decreasing with $l^{(0)} = 12$ and $l^* = 6$, the corresponding thresholds are $a = 4.265, b = 0.661$, and $c = 0.665$, respectively.

Figure 11 illustrates the detection delay ess sup $\mathbf{E}_\nu\big((T - \nu + 1)^+ \big| \mathcal{F}_{\nu-1}\big)$ at different change-points ν for each of the three detection schemes. When the step function of population sizes is increasing (the left panel of Figure 11), it is interesting to note that for

the GLR scheme $T_{GLR}(a)$, the detection delay ess sup $\mathbf{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$ is a decreasing function of change point ν (decreased from 36.9 ± 0.1 at $\nu = 0$ to 19.1 ± 0.1 at $\nu = 200$). However, the (simulated) detection delay of either $T_{WLR}(b)$ or $T_{ATM}(c)$ is an increasing function of change point ν (increased from 20.4 ± 0.1 at $\nu = 0$ to 23.1 ± 0.1 at $\nu = 200$). Hence, under Lorden's worst-case detection delay criterion, the worst-case detection delays $\overline{\mathbf{E}}_1(T)$ of $T_{GLR}(a), T_{WLR}(b)$, and $T_{ATM}(c)$ are $36.9 \pm 0.1, 23.1 \pm 0.1$, and 23.1 ± 0.1, respectively. Thus, if the step function of the population sizes is increasing, the schemes $T_{WLR}(b)$ and $T_{ATM}(c)$ are better than the scheme $T_{GLR}(a)$ in the sense that they have smaller detection delays under Lorden's worst-case detection delay criterion. This is consistent with the asymptotic theory developed in Section 4.7.1 for the case of increasing population sizes with $l^{(0)} < l^*$.

On the other hand, if the step function of the population size is decreasing (the right panel of Figure 11), we get the reverse pattern: the detection delay ess sup $\mathbf{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$ is an increasing function of change point ν for $T_{GLR}(a)$, but it is a decreasing function of ν for either $T_{WLR}(b)$ or $T_{ATM}(c)$. Under Lorden's worst-case detection delay criterion, the worst-case detection delays $\overline{E}(T)$ of $T_{GLR}(a), T_{WLR}(b)$, and $T_{ATM}(c)$ are $34.4 \pm 0.1, 35.0 \pm 0.1$, and 34.7 ± 0.1, respectively. Hence, when the step function of the population sizes is decreasing, despite significantly different (individual) detection delay curves illustrated in the right panel of Figure 11, the three proposed schemes have similar properties under Lorden's criterion! This finding is also consistent with the asymptotic theory developed in Section 4.7.1 with the case of decreasing population sizes $l^{(0)} > l^*$.

Finally, Figure 11 shows that compared to those for the GLR scheme, the detection delay curves for $T = T_{WLR}(b)$ or $T_{ATM}(c)$ appear to be flatter with respect to change point ν, regardless of whether the population sizes are increasing or decreasing. This finding supports our claim that the GLR scheme $T_{GLR}(a)$ is not an "equalizer rule", even at the first-order, whereas the other two alternative schemes, $T_{WLR}(b)$ and $T_{ATM}(c)$, are (first-order) "equalizer rules."

4.8 Conclusion

In this chapter we have studied the problem of detecting a change in the mean of Poisson distributions after taking into account the effect of population sizes. Such a problem has an important application in biosurveillance when one is interested in monitoring whether the disease risk changes or not. Despite its asymptotic optimality properties under the classical asymptotic setting, the GLR scheme can have poor finite sample properties as compared to the two alternative ad-hoc schemes: the weighted likelihood ratio (WLR) and the adaptive threshold method (ATM). To understand why the GLR is at times efficient and at times inefficient, we have studied a new asymptotic setting by assuming that the time at which the population sizes reach the stationary value is comparable to the value in the false alarm constraint. Asymptotic properties of the three schemes are established under the new asymptotic setting, and they are consistent with our finite-sample numerical simulations.

4.9 Appendix: Proof of Mathematical Results

4.9.1 Proof of Theorem 4.1

Define $Z_i = \log \frac{f_{\lambda_1}(Y_i)}{f_{\lambda_0}(Y_i)}$, the log-likelihood ratio statistic, and $I = l^* I(\lambda_1, \lambda_0) > 0$. By Theorem 1 of Lai [55] (page 2919), Theorem 4.1 holds if

$$\lim_{n\to\infty} \sup_{\nu\geq 1} P^{(\nu)}\Big\{\frac{1}{n}\max_{1\leq t\leq n}\sum_{i=\nu+1}^{\nu+t} Z_i \geq I+\delta\Big\} = 0 \text{ for any } \delta > 0. \tag{31}$$

To prove (31), note that

$$\begin{aligned}
& P^{(\nu)}\Big\{\frac{1}{n}\max_{1\leq t\leq n}\sum_{i=\nu+1}^{\nu+t} Z_i \geq I+\delta\Big\} \\
= \; & P^{(\nu)}\Big\{\frac{1}{n}\max_{1\leq t\leq n}\sum_{i=\nu+1}^{\nu+t} \Big(Z_i - E[Z_i] + E[Z_i]\Big) \geq I+\delta\Big\} \\
\leq \; & P^{(\nu)}\Big\{\frac{1}{n}\max_{1\leq t\leq n}\sum_{i=\nu+1}^{\nu+t} \Big(Z_i - E[Z_i]\Big) + \sup_{\nu}\frac{1}{n}\sum_{i=\nu+1}^{\nu+t} l_i I(\lambda_1,\lambda_0) \geq I+\delta\Big\} \qquad (32)
\end{aligned}$$

since $l_i > 0$. By the uniform assumption (22), given δ, there exists N such that for $n \geq N$,

$$\sup_{\nu}\frac{1}{n}\sum_{\nu+1}^{\nu+t} l_i \leq l^* + \frac{\delta}{2I(\lambda_1,\lambda_0)}.$$

Thus, from Inequality (32), for $n \geq N$,

$$P^{(\nu)}\Big\{\frac{1}{n}\max_{1\leq t\leq n}\sum_{i=\nu+1}^{\nu+t} Z_i \geq I+\delta\Big\} \leq P^{(\nu)}\Big\{\frac{1}{n}\max_{1\leq t\leq n}\sum_{i=\nu+1}^{\nu+t} R_i \geq \delta/2\Big\},$$

where $R_i = Z_i - E[Z_i]$.

By Skorohod's inequality (Resnick [79], p. 209), for any δ,

$$\begin{aligned} & P^{(\nu)}\Big(|\frac{1}{n}\sum_{i=\nu+1}^{\nu+n} R_i| > \frac{\delta}{4}\Big) \\ \geq\ & P^{(\nu)}\Big(\frac{1}{n}\max_{1\leq t\leq n}|\sum_{i=\nu+1}^{\nu+t} R_i| > \frac{\delta}{2}\Big) \times \min_{1\leq k\leq n} P^{(\nu)}\Big(\frac{1}{n}|\sum_{i=\nu+k}^{\nu+n} R_i| < \frac{\delta}{4}\Big) \end{aligned}$$

Thus, it suffices to show that

$$\lim_{n\to\infty}\sup_{\nu} P^{(\nu)}\Big(|\frac{1}{n}\sum_{i=\nu+1}^{\nu+n} R_i| > \frac{\delta}{4}\Big) = 0 \tag{33}$$

and

$$\lim_{n\to\infty}\inf_{\nu}\min_{1\leq k\leq n} P^{(\nu)}\Big(\frac{1}{n}|\sum_{i=\nu+k}^{\nu+n} R_i| < \frac{\delta}{4}\Big) = 1. \tag{34}$$

First, let us prove Equation (33). By Chebyshev's inequality, $P^{(\nu)}\Big\{\Big|\frac{1}{n}\sum_{i=\nu+1}^{\nu+n} R_i\Big| > \frac{\delta}{4}\Big\}$ is bounded by $\frac{1}{(\delta/4)^2 n}\sup_i Var(R_i)$ for all $\nu \geq 1$, and the bound does not depend on ν. Thus, (33) is satisfied.

Next, let us prove Equation (34). By Chebyshev's inequality, for all k such that $1 \leq k \leq n$,

$$\begin{aligned} P^{(\nu)}\{|\frac{1}{n}\sum_{i=\nu+k}^{\nu+n} R_i| \geq \frac{\delta}{4}\} &\leq \frac{1}{(\delta/4)^2 n^2} \times \Big(\sum_{i=\nu+k}^{\nu+n} Var(R_i)\Big) \\ &\leq \frac{1}{(\delta/4)^2 n} \times \sup_i Var(R_i). \end{aligned}$$

Letting $n \to \infty$ yields relation 34, as the upper bound does not depend n or ν.

4.9.2 Proof of Theorem 4.2

Theorem 4.4 (i) for $E_\infty(T_{GLR}(a))$ follows directly from Theorem 4 of Lai [55] (page 2921). To prove Theorem 4.2 (ii), it suffices to show that under the uniform convergence assumption of (22), the detection delay $\bar{E}_1(T_{GLR}(a)) \leq \frac{(1+o(1))a}{l^* I(\lambda_1,\lambda_0)}$.

Define $S_n = \sum_{i=1}^n Z_i$ for $n \geq 1$ and $S_0 = 0$, where $Z_i = \log \frac{f_{\lambda_1}(Y_i)}{f_{\lambda_0}(Y_i)}$. Consider the open-ended test $\tau = \inf\left\{n \geq 1 : \sum_{i=1}^n Z_i \geq a\right\}$. Following the ideas in Lorden [57] and Yao [114], denote by τ_ν the stopping time obtained by applying τ to $Y_\nu, Y_{\nu+1}, ...$, i.e. $\tau_\nu = \inf\left\{i \geq 1 : \sum_{j=\nu}^{\nu+i-1} Z_j \geq a\right\}$. Then, by $T_{GLR}(a) = \inf_{\nu\geq 1}\left\{\tau_\nu + \nu - 1\right\}$, $E_\nu\Big[(T_{GLR}(a) - \nu + 1)^+\Big|Y_1, Y_2, ..., Y_{\nu-1}\Big] \leq E_\nu[\tau_\nu]$, and thus, $\bar{E}_1[T_{GLR}(a)] \leq \sup_{\nu\geq 1} E_\nu[\tau_\nu]$.

It suffices to show that for all ν, $E_\nu[\tau_\nu]$ is bounded above by $\frac{(1+o(1))a}{l^* I(\lambda_1,\lambda_0)}$, where $o(1)$ does not depend on ν. Fix a ν. By assumption (22),

$$\frac{1}{n}\sum_{i=\nu}^{\nu-n+1} l_i I(\lambda_1,\lambda_0) \to l^* I(\lambda_1,\lambda_0) \text{ as } n \to \infty.$$

Furthermore, by the uniformity of the convergence of (22), $\sup_{i\geq 1} E_1[Z_i] = \sup_{i\geq 1} l_i I(\lambda_1,\lambda_0) < \infty$, and there exists a positive integer d such that $\frac{1}{2} l^* I(\lambda_1,\lambda_0) < \frac{1}{d}\sum_{i=k+1}^{k+d} l_i I(\lambda_1,\lambda_0) < \frac{3}{2} l^* I(\lambda_1,\lambda_0)$ for all $k \geq 0$.

Let $S_0 = 0$. $\tilde{Z}_n = S_{\nu+nd} - S_{\nu+(n-1)d} = \sum_{i=\nu+(n-1)d+1}^{\nu+nd} Z_i$, $n \geq 1$. Then, under P_ν, the mean of $\tilde{Z}_n$ is $\sum_{i=\nu+(n-1)d+1}^{\nu+(n-1)d+d} l_i I(\lambda_1,\lambda_0)$, where $\frac{dl^* I(\lambda_1,\lambda_0)}{2} \leq E[\tilde{Z}_n] \leq \frac{3dl^* I(\lambda_1,\lambda_0)}{2}$. Let $t_\nu(a) = \inf\{i \geq 1 | \sum_{j=1}^i \tilde{Z}_j \geq a\}$. Obviously, $\tau_\nu(a) \leq d t_\nu(a)$.

The key part of the proof in this subsection is to show that there exists a positive K to satisfy that $\min_{n\geq 1} E_n[W_n] \geq \epsilon$, where $W_n = \min\{\tilde{Z}_n, K\log(\lambda_1/\lambda_0) - \tilde{l}_n(\lambda_1 - \lambda_0)\}$, and ϵ is a positive constant. To prove this, we decompose $E[W_n]$ by the following,

$$E_\nu[W_n] \;\; = \;\; E_\nu[\tilde{Z}_n - \tilde{Z}_n + W_n] = E_\nu[\tilde{Z}_n] - E_\nu[\tilde{Z}_n - W_n].$$

Let $X_n = \sum_{i=\nu+(n-1)d+1}^{\nu+nd} Y_i$. If $\lambda_1 > \lambda_0$, then

$$\begin{aligned} E_\nu[\tilde{Z}_n - W_n] &= E_\nu[\tilde{Z}_n - W_n; \tilde{Z}_n \geq K\log(\lambda_1/\lambda_0) - \tilde{l}_n(\lambda_1 - \lambda_0)] \\ &= E_\nu[\log(\lambda_1/\lambda_0)(X_n - K); X_n \geq K] \\ &= \sum_{X_n=K}^{\infty} \log(\lambda_1/\lambda_0)(X_n - K)\frac{(\tilde{l}_n\lambda_1)^{X_n}e^{-\tilde{l}_n\lambda_1}}{X_n!} \\ &= (\tilde{l}_n\lambda_1)\log(\lambda_1/\lambda_0)P(X_n \geq K-1) - \log(\lambda_1/\lambda_0)KP(X_n \geq K) \\ &\leq (\tilde{l}_n\lambda_1)\log(\lambda_1/\lambda_0)P(X_n \geq K-1). \end{aligned}$$

For ϵ such that $0 < \epsilon < \tilde{l}_n I(\lambda_1, \lambda_0)$, there exists K to satisfy that

$$\tilde{l}_n\lambda_1\log(\lambda_1/\lambda_0)P(X_n \geq K-1) \leq \tilde{l}_n I(\lambda_1, \lambda_0) - \epsilon,$$

which indicates that for K, $E_\nu[W_n] = \tilde{l}_n I(\lambda_1, \lambda_0) - E_\nu[\tilde{Z}_n - W_n] > \epsilon > 0$. Thus,

$$E_\nu[W_n] > \epsilon.$$

Similarly, if $\lambda_1 < \lambda_0$, such a K also exists.

For both cases of $\lambda_0 < \lambda_1$ and $\lambda_0 > \lambda_1$,$\min_{n\geq 1} E_\nu[W_n] > 0$, and $K\log(\lambda_1/\lambda_0) - \tilde{l}_n(\lambda_1 - \lambda_0) \leq K\log(\lambda_1/\lambda_0) - \frac{3dl^*}{2}(\lambda_1 - \lambda_0) < \infty$. Consequently, by Lemma 3 (iii) in Yao [114], if

$$\min_{n\geq 1} E_\nu[\min\{\tilde{Z}_n, K\log(\lambda_1/\lambda_0) - \tilde{l}_n(\lambda_1 - \lambda_0)\}] > 0$$

for $K\log(\lambda_1/\lambda_0) - \tilde{l}_n(\lambda_1 - \lambda_0) < \infty$, then $E_\nu[t_\nu(a)] < K_2 < \infty$ for all a with

$$K_2 = \frac{a + \max_{n\geq 1} K\log(\lambda_1/\lambda_0) - \tilde{l}_n(\lambda_1 - \lambda_0)}{\min_{n\geq 1} E[\min\{Z_n, K\log(\lambda_1/\lambda_0) - \tilde{l}_n(\lambda_1 - \lambda_0)\}]} > 0.$$

Furthermore, $E_\nu[\tau_\nu(a)] \leq dE_\nu[t_\nu(a)] \leq dK_2 < \infty$ for all a, and the bound dK_2 does not depend on ν. For fixed ν, it is easy to show

$$a \leq E_\nu[\sum_{i=\nu}^{\nu+\tau_\nu(a)-1} Z_i] \leq a + \sup_{i\geq 1} E[|Z_i|], \tag{35}$$

where $\sup_{i\geq 1} E[|Z_i|]$ does not depend on ν. By the uniform convergence assumption of (22) and the monotonicity of l_n,

$$\frac{E_\nu[\sum_{i=\nu}^{\nu+\tau_\nu(a)-1} Z_i]}{E_\nu[\tau_\nu(a)]} \to l^* I(\lambda_1, \lambda_0), \tag{36}$$

which dose not depends on ν. It is proven in Lemma 4.1. Therefore, from Inequality (35) and Relation (36), $\frac{E_\nu[\tau_\nu(a)]}{a} \to \frac{1}{l^* I(\lambda_1,\lambda_0)}$ uniformly by ν. ■

Lemma **4.1** *By the uniform convergence assumption of (22) and the monotonicity of l_n,*

$$\frac{E_\nu[\sum_{i=\nu}^{\nu+\tau_\nu(a)-1} Z_i]}{E_\nu[\tau_\nu(a)]} \to l^* I(\lambda_1, \lambda_0),$$

which dose not depends on ν.

Proof.

For fixed a, let $H^a_{\nu,i} = 1$ if $\tau_\nu(a) \geq i$, 0 if $\tau(a) < i$.

$$\begin{aligned} E_\nu[\sum_{i=\nu}^{\nu+\tau_\nu(a)-1} Z_i] &= \sum_{i=\nu}^{\infty} E[Z_i H^a_{\nu,i}] \\ &= \sum_{i=\nu}^{\infty} E[Z_i] E[H^a_{\nu,i}] \\ &= \sum_{i=\nu}^{\infty} \{(E[Z_i] - l^* I(\lambda_1,\lambda_0)) E[H^a_{\nu,i}] + l^* I(\lambda_1,\lambda_0) E[H^a_{\nu,i}]\} \\ &= \sum_{i=\nu}^{\infty} (E[Z_i] - l^* I(\lambda_1,\lambda_0)) E[H^a_{\nu,i}] + l^* I(\lambda_1,\lambda_0) \sum_{i=\nu}^{\infty} E[H^a_{\nu,i}]. \end{aligned}$$

Thus, for the given a,

$$\frac{E_\nu[\sum_{i=\nu}^{\nu+\tau_\nu(a)-1} Z_i]}{\sum_{i=\nu}^{\infty} E[H^a_{\nu,i}]} = \frac{\sum_{i=\nu}^{\infty} (E[Z_i] - l^* I(\lambda_1,\lambda_0)) E[H^a_{\nu,i}]}{\sum_{i=\nu}^{\infty} E[H^a_{\nu,i}]} + l^* I(\lambda_1,\lambda_0).$$

Since $\sum_{i=\nu}^{\infty} E[H^a_{\nu,i}] = E[\tau_\nu(a)]$,

$$\sup_\nu \Big| \frac{E_\nu[\sum_{i=\nu}^{\nu+\tau_\nu(a)-1} Z_i]}{E[\tau_\nu(a)]} - l^* I(\lambda_1,\lambda_0) \Big| = \sup_\nu \Big| \frac{\sum_{i=\nu}^{\infty} (E[Z_i] - l^* I(\lambda_1,\lambda_0)) E[H^a_{\nu,i}]}{\sum_{i=\nu}^{\infty} E[H^a_{\nu,i}]} \Big|. \quad (37)$$

In addition, by the uniform convergence assumption of (22) and the monotonicity of l_n, $l_i \to l^*$, which is defined by that

$$\text{for all } \varepsilon > 0, \text{ there exists } N(\varepsilon) \text{ such that } i \geq N(\varepsilon) \Rightarrow |l_i - l^*| < \varepsilon. \quad (38)$$

Thus, from (37) and (38), for any fixed ε and any a,

$$\sup_\nu \Big| \frac{\sum_{i=\nu}^{\infty} (E[Z_i] - l^* I(\lambda_1,\lambda_0)) E[H^a_{\nu,i}]}{\sum_{i=\nu}^{\infty} E[H^a_{\nu,i}]} \Big| \leq \varepsilon I(\lambda_1,\lambda_0) + \frac{N(\varepsilon) \sup_i |l_i - l^*| I(\lambda_1,\lambda_0)}{E[\tau_\nu(a)]},$$

which indicates that for any fixed ε,

$$\lim_{a\to\infty} \sup_{\nu} \left| \frac{\sum_{i=\nu}^{\infty} (E[Z_i] - l^* I(\lambda_1, \lambda_0)) E[H_{\nu,i}^a]}{\sum_{i=\nu}^{\infty} E[H_{\nu,i}^a]} \right| \le \varepsilon I(\lambda_1, \lambda_0).$$

Therefore,

$$\lim_{a\to\infty} \sup_{\nu} \left| \frac{\sum_{i=\nu}^{\infty} (E[Z_i] - l^* I(\lambda_1, \lambda_0)) E[H_{\nu,i}^a]}{\sum_{i=\nu}^{\infty} E[H_{\nu,i}^a]} \right| = 0,$$

and by (37),

$$\lim_{a\to\infty} \sup_{\nu} \left| \frac{E_\nu[\sum_{i=\nu}^{\nu+\tau_\nu(a)-1} Z_i]}{E[\tau_\nu(a)]} - l^* I(\lambda_1, \lambda_0) \right| = 0.$$

4.9.3 Proof of Theorem 4.3

Under the assumption of Theorem 4.3, the lower bound (23) can be established by using the arguments in either Lorden [57] or Lai [55]. Here, let us follow the proof of Theorem 1 in Lai [55] with a very minor twist. Let m be a positive integer $< \gamma - \omega$. A key observation is that if $\mathbf{E}_\infty(T) \ge \gamma$, then for some $\nu \ge \omega$

$$\mathbf{P}_\infty(T \ge \nu) > 0 \quad \text{and} \quad \mathbf{P}_\infty(T < \nu + m | T \ge \nu) \le \frac{m}{\gamma - \omega}. \tag{39}$$

Otherwise, $\mathbf{P}_\infty(T \ge \nu + m | T \ge \nu) < 1 - \frac{m}{\gamma-\omega}$ for all $\nu \ge \omega$ with $\mathbf{P}_\infty(T \ge \nu) > 0$, implying that

$$\begin{aligned}
\mathbf{E}_\infty(T) &= \sum_{k=1}^{\infty} \mathbf{P}_\infty(T \ge k) \\
&= \sum_{k=1}^{\omega} \mathbf{P}_\infty(T \ge k) + \sum_{k=1}^{\infty} \mathbf{P}_\infty(T \ge \omega + k) \\
&= \sum_{k=1}^{\omega} \mathbf{P}_\infty(T \ge k) + \sum_{i=1}^{m} \sum_{j=0}^{\infty} \mathbf{P}_\infty(T \ge \omega + i + mj) \\
&< \omega + \sum_{i=1}^{m} \sum_{j=0}^{\infty} \Big(1 - \frac{m}{\gamma - \omega}\Big)^j \mathbf{P}_\infty(T \ge \omega + i) \\
&\le \omega + (\gamma - \omega) = \gamma.
\end{aligned}$$

Now the observations Y_n's are i.i.d. Poisson distributed with constant population size l^* for all $n \ge \nu$. Following the arguments in Lai [55], let m be the largest integer $\le (\log(\gamma - \omega))^2$ for such $\nu \ge \omega$, satisfying (39).

For all $\delta > 0$,

$$\sup_{\nu \geq 1} E_\nu \Big[T - \nu \Big| T \geq \nu \Big] \geq P^{(\nu)} \Big\{ T - \nu \geq (1-\delta) \frac{\log(\gamma - \omega)}{I} \Big| T \geq \nu \Big\} \times (1-\delta) \frac{\log(\gamma - \omega)}{I}, \tag{40}$$

where $I = l^* I(\lambda_1, \lambda_0)$. Therefore, to prove Theorem 4.3, it suffices to show that for any $\delta > 0$, as $(\gamma - \omega) \to \infty$,

$$P^{(\nu)} \Big\{ T - \nu \geq (1-\delta) \frac{\log(\gamma - \omega)}{I} \Big| T \geq \nu \Big\} \longrightarrow 1,$$

or equivalently,

$$P^{(\nu)} \Big\{ 0 \leq T - \nu < (1-\delta) \frac{\log(\gamma - \omega)}{I} \Big| T \geq \nu \Big\} \longrightarrow 0. \tag{41}$$

One way of proving (41) is to take two disjoint sets A_δ and B_δ, where

$$A_\delta = \{ 0 \leq T - \nu < (1-\delta) \frac{\log(\gamma - \omega)}{I}, \sum_{i=\nu}^{T} Z_i \geq (1 - \delta^2) \log(\gamma - \omega) \},$$

and

$$B_\delta = \{ 0 \leq T - \nu < (1-\delta) \frac{\log(\gamma - \omega)}{I}, \sum_{i=\nu}^{T} Z_i < (1 - \delta^2) \log(\gamma - \omega) \}.$$

From the proof in Lai [55], $P\Big(A_\delta \Big| T \geq \nu \Big)$ and $P\Big(B_\delta \Big| T \geq \nu \Big)$ go to 0; then Equation (41) is satisfied. Thus, from Equation (40),

$$\sup_{\nu \geq 1} \mathbf{E}_\nu (T - \nu | T \geq \nu) \geq (1 + o(1)) \frac{\log(\gamma - \omega)}{l^* I(\lambda_1, \lambda_0)}.$$

Theorem 4.3 follows directly from this relation and the fact that $\log(\gamma - \omega) = \log \gamma + O(1)$ when $0 < \omega < C\gamma$ for some constant $0 < C < 1$ since $\log(\gamma - \omega) \geq \log(\gamma - C\gamma) = \log \gamma + \log(1 - C)$, where the term $\log(1 - C)$ is a constant.

4.9.4 Proof of Theorem 4.4

Define $S_n = \sum_{i=1}^{n} Z_i$ for $n \geq 1$ and $S_0 = 0$, where $Z_i = \log \frac{f_{\lambda_1}(Y_i)}{f_{\lambda_0}(Y_i)}$. Consider the open-ended test

$$\tau = \inf \big\{ n \geq 1 : \sum_{i=1}^{n} Z_i \geq b \big\}.$$

Following the ideas in Lorden [57], denote by τ_ν the stopping time obtained by applying τ to $Y_\nu, Y_{\nu+1}, \ldots$. Then, $T_{WLR}(b) = \inf_{\nu \geq 1}\{\tau_\nu + \nu - 1\}$. For any change time ν, $T_{WLR}(b) \leq \tau_\nu + \nu + 1$. Thus,

$$\bar{E}_1\Big[T_{WLR}(b)\Big] = \sup_{1\leq\nu} ess\sup E\Big[(T_{WLR}(b)-\nu+1)^+\Big|Y_1, Y_2, ..., Y_{\nu-1}\Big] \leq \sup_{1\leq\nu} E_\nu[\tau_\nu]. \quad (42)$$

To prove this, we focus on the open-ended test τ_ν under the probability measure $\mathbf{P}_\nu$. To simplify the notation, in the following proof, we simply use $\mathbf{E}_\nu$ to denote the expectation under $\mathbf{P}_\nu$. Let $V = \lambda_1(\log(\lambda_1/\lambda_0))^2$. It is easy to see that $\mathbf{E}(Z_i) = I(\lambda_1, \lambda_0) > 0$ and $Var(Z_i) = V/l_i$ is uniformly bounded by V/l_*. Hence $\tau_\nu < \infty$ a.s. By a modification of Wald's equation (on the independent random variables with the same mean), $\mathbf{E}_\nu(S_{\tau_\nu}) = I(\lambda_1, \lambda_0)\mathbf{E}(\tau_\nu)$. On the one hand, by the definition of τ_ν, we have $S_{\tau_\nu} \geq b$, and thus, $\mathbf{E}(\tau_\nu) \geq b/I(\lambda_1, \lambda_0)$. On the other hand, using the fact that $S_{\tau_\nu - 1} < b$, we have

$$\begin{aligned}
I(\lambda_1, \lambda_0)\mathbf{E}_\nu(\tau_\nu) &= \mathbf{E}_\nu(S_{\tau_\nu}) = \mathbf{E}_\nu(S_{\tau_\nu-1} + Z_{\tau_\nu}) < b + \mathbf{E}_\nu|Z_{\tau_\nu}| \\
&\leq b + \sup_{i\geq\nu} \mathbf{E}_\nu|Z_i| \\
&\leq b + \sup_{i\geq\nu} \sqrt{\mathbf{E}_\nu(Z_i^2)} = b + \sup_{i\geq\nu} \sqrt{(\mathbf{E}_\nu Z_i)^2 + Var(Z_i)} \\
&= b + \sup_{i\geq1} \sqrt{(I(\lambda_1, \lambda_0))^2 + V/l_i} \\
&= b + \sqrt{(I(\lambda_1, \lambda_0))^2 + V/l_*}.
\end{aligned}$$

Hence, a simple calculation shows that for all ν,

$$\frac{b}{I(\lambda_1, \lambda_0)} \leq \mathbf{E}_\nu(\tau_\nu) \leq \frac{b}{I(\lambda_1, \lambda_0)} + M,$$

where M is the constant given in the theorem. The theorem then follows directly from the fact that the above relation holds uniformly for all possible subsets of population sizes l_n's, which are uniformly bounded below by l_*. ■

4.9.5 Proof of Theorem 4.5

First, it is well known that the detection delays are on the order of $\log\gamma$ subject to the false alarm constraint in (19). Since $\log\gamma << \omega$, for each of the three proposed schemes, the

Table 5: Detection delays of the three proposed detection schemes at change-point $\nu = 1$ or ω

detection scheme T	$\mathbf{E}_1(T)$	ess sup $\mathbf{E}_\omega\Big((T-\omega+1)^+ \vert \mathcal{F}_{\omega-1}\Big)$
$T_{GLR}(a)$	$(1+o(1))\frac{a}{l^{(0)}I(\lambda_1,\lambda_0)}$	$(1+o(1))\frac{a}{l^*I(\lambda_1,\lambda_0)}$
$T_{WLR}(b)$	$(1+o(1))b/I(\lambda_1,\lambda_0)$	$(1+o(1))b/I(\lambda_1,\lambda_0)$
$T_{ATM}(c)$	$(1+o(1))c/I(\lambda_1,\lambda_0)$	$(1+o(1))c/I(\lambda_1,\lambda_0)$

detection delay $\mathbf{E}_{\nu=1}(T)$ when the change point occurring at $\nu = 1$ is mainly determined by the first stage when population size $l_n = l^{(0)}$. Thus, all three schemes become Page's CUSUM procedure with constant population sizes $l^{(0)}$, and the classical results can be applied to determine $\mathbf{E}_{\nu=1}(T)$. On the other hand, when a change occurs at time ω, all incoming new post-change observations have population size l^*, and thus, all three schemes become Page's CUSUM procedure with constant population sizes l^*. Hence, the detection delay ess sup $\mathbf{E}_\omega\big((T-\omega+1)^+|\mathcal{F}_{\omega-1}\big)$ can also be easily calculated by the classical results since it is just the detection delay in the classical change-point detection problem with i.i.d. models in which one detects a change in the Poisson mean from $l^*\lambda_0$ to $l^*\lambda_1$. Table 5 shows the detection delays of the three proposed schemes when a change occurs at time $\nu = 1$ or ω as thresholds a, b, and c go to ∞.

Next, we need to establish the relationship between the false alarm constraint γ in (19) and the threshold values a, b, and c in the three proposed detection schemes $T_{GLR}(a), T_{WLR}(b)$, and $T_{ATM}(c)$, respectively. Intuitively, when $\omega < C\gamma$ for some $0 < C < 1$, we have to take at least $\gamma - \omega$ observations from the stage with population sizes $l_n = l^*$ in order to satisfy the false alarm constraint. Now at the stage with $l_n = l^*$, all three proposed detection schemes become the classical CUSUM procedure under the i.i.d. model. Thus, intuitively, the classical results imply that $a \sim \log\gamma, b \sim \log\gamma/l^*$, and $c \sim \log\gamma/l^*$. By combining the above results, we complete the proof of Theorem 4.5.

Now, let us present a rigorous proof that $b \sim \log\gamma/l^*$. The proofs of $a \sim \log\gamma$ and $c \sim \log\gamma/l^*$ are similar. The key idea is to place a condition on $E_\infty[T_{WLR}(b)]$ by population

change time ω such that

$$\begin{aligned} E_\infty[T_{WLR}(b)] &= E_\infty[T_{WLR}(b)|T_{WLR}(b) < \omega]P(T_{WLR}(b) < \omega) \quad (43) \\ &\quad + E_\infty[T_{WLR}(b)|T_{WLR}(b) \geq \omega]P(T_{WLR}(b) \geq \omega) \\ &= E_\infty[T_{WLR}(b)|T_{WLR}(b) \leq \omega]P(T_{WLR}(b) \leq \omega) \\ &\quad + \int_0^b E_\infty[\omega + T^*_{Cusum}(l^*\tilde{W}_\omega, l^*b)]d_{F(\tilde{W}_\omega)}P(T_{WLR}(b) > \omega), \quad (44) \end{aligned}$$

where $T^*_{Cusum}(z, l^*b)$ be a detection scheme of CUSUM with initial value z and threshold l^*b under constant population l^*. In other words, $T^*_{Cusum}(z, l^*b) = \inf\left\{n \geq 1 | \tilde{C}_n \geq l^*b\right\}$, where $\tilde{C}_0 = z$, and $\tilde{C}_n = \max(0, \tilde{C}_{n-1} + \log \frac{f_{l^*\lambda_1}(Y_n)}{f_{l^*\lambda_0}(Y_n)})$. In addition, $F(\cdot)$ is the distribution of statistics $\tilde{W}_\omega$ at time ω, where $\tilde{W}_n = \max\left(0, \tilde{W}_{n-1} + \frac{1}{l^{(0)}} \log \frac{f_{l^{(0)}\lambda_1}(Y_n)}{f_{l^{(0)}\lambda_0}(Y_n)}\right)$ and $\tilde{W}_0 = 0$.

First, let us show that $\log \gamma \leq (1 + o(1))l^*b$. If Equation (44) holds, then

$$\begin{aligned} E_\infty[T_{WLR}(b)] &\leq \omega P(T_{WLR}(b) \leq \omega) \\ &\quad + \int_0^b E_\infty[\omega + T^*_{Cusum}(l^*\tilde{W}_\omega, l^*b)]d_{F(\tilde{W}_\omega)}P(T_{WLR}(b) > \omega) \\ &\leq \omega + \int_0^b E_\infty[T^*_{Cusum}(l^*\tilde{W}_\omega, l^*b)]d_{F(\tilde{W}_\omega)}P(T_{WLR}(b) > \omega) \\ &\leq \omega + E_\infty[T^*_{Cusum}(0, l^*b)] = \omega + D(1 + o(1))e^{l^*b}, \end{aligned}$$

where D is some constant. If we want $E_\infty[T_{WLR}(b)] \geq \gamma$, then we must have

$$\log \gamma \leq \log(1 + o(1)) + l^*b, \text{, so } b \geq \frac{\log \gamma + o(1)}{l^*}. \quad (45)$$

Second, let us prove the other direction $\log E_\infty[T_{WLR}(b)] \geq l^*b + o(1)$. Ignoring the first term in Equation (44), we have

$$\begin{aligned} E_\infty[T_{WLR}(b)] &\geq \int_0^b \omega + E_\infty[T^*_{Cusum}(l^*\tilde{W}_\omega, l^*b)]d_{F(\tilde{W}_\omega)}P(T_{WLR}(b) > \omega) \\ &\geq \int_0^b E_\infty[T^*_{Cusum}(l^*\tilde{W}_\omega, l^*b)]d_{F(\tilde{W}_\omega)}P(T_{WLR}(b) > \omega) \\ &\geq \int_0^{\eta b} E_\infty[T^*_{Cusum}(l^*\tilde{W}_\omega, l^*b)]d_{F(\tilde{W}_\omega)}P(T_{WLR}(b) > \omega), \end{aligned}$$

where η is defined in Theorem 4.5. Since $E_\infty[T^*_{Cusum}(z, l^*b)]$ is decreasing as a function of z, we have

$$\begin{aligned} E_\infty[T_{WLR}(b)] &\geq \int_0^{\eta b} E_\infty[T^*_{Cusum}(l^*\eta b, l^*b)] d_{F(\tilde{W}_\omega)} P(T_{WLR}(b) > \omega) \\ &= E_\infty[T^*_{Cusum}(l^*\eta b, l^*b)] \int_0^{\eta b} d_{F(\tilde{W}_\omega)} P(T_{WLR}(b) > \omega) \\ &= E_\infty[T^*_{Cusum}(l^*\eta b, l^*b)] F(\eta b) P(T_{WLR}(b) > \omega). \end{aligned}$$

Taking the logarithm on both sides,

$$\log E_\infty[T_{WLR}(b)] \geq \log E_\infty[T^*_{Cusum}(l^*\eta b, l^*b)] + \log\Big(F(\eta b) P(T_{WLR}(b) \geq \omega)\Big).$$

Each term on the right-hand side of the equation is bounded below by the following two lemmas, so $(\log E_\infty[T_{WLR}(b)]) \geq (1 + o(1))l^*b + \log(1 + o(1))$. This completes the proof that $b \sim \log\gamma/l^*$. ■

We still need to prove Lemmas 4.2 and 4.3.

Lemma **4.2** $\log E_\infty[T^*_{Cusum}(l^*\eta b, l^*b)] \geq (1 + o(1))l^*b$.

Proof. The main idea is that the Shiryaev-Roberts procedure ($T_{SR}(l^*b)$) produces a lower bound on $T_{Cusum}(l^*b)$, and $T_{SR}(l^*b)$ has a martingale property. The Shiryaev-Roberts procedure ($T_{SR}(l^*b)$) is defined by

$$T_{SR}(l^*b) = \inf\{n \geq 1 | \tilde{R}_n \geq e^{l^*b}\},$$

where $\tilde{R}_0 = e^z$, and $\tilde{R}_n = (1 + \tilde{R}_{n-1})\frac{f_{l^*\lambda_1}(Y_n)}{f_{l^*\lambda_0}(Y_n)}$. $\tilde{R}_n$ is the Shiryaev-Roberts statistic given by

$$\tilde{R}_n = \sum_{\nu=0}^{n} \frac{f_{l^*\lambda_1}(Y_\nu) \cdots f_{l^*\lambda_1}(Y_n)}{f_{l^*\lambda_0}(Y_\nu) \cdots f_{l^*\lambda_0}(Y_n)},$$

where $\frac{f_{l^*\lambda_1}(Y_0)}{f_{l^*\lambda_0}(Y_0)} = z$. $\tilde{C}_n$ denotes the Cusum statistic before, and

$$e^{\tilde{C}_n} = \max_{0 \leq \nu \leq n} \frac{f_{l^*\lambda_1}(Y_\nu) \cdots f_{l^*\lambda_1}(Y_n)}{f_{l^*\lambda_0}(Y_\nu) \cdots f_{l^*\lambda_0}(Y_n)}.$$

First, let us show that $\tilde{R}_n \geq e^{\tilde{C}_n}$. If $n = 0$, then it is true. Assume that when $n = k$, the inequality is true. If $n = k + 1$, then

$$\tilde{R}_{k+1} = (1+\tilde{R}_k)\frac{f_{l^*\lambda_1}(Y_{k+1})}{f_{l^*\lambda_0}(Y_{k+1})} \geq \max(1, \tilde{R}_k)\frac{f_{l^*\lambda_1}(Y_{k+1})}{f_{l^*\lambda_0}(Y_{k+1})} \geq \max(1, e^{\tilde{C}_k})\frac{f_{l^*\lambda_1}(Y_{k+1})}{f_{l^*\lambda_0}(Y_{k+1})} = e^{\tilde{C}_{k+1}}.$$

Second, let's prove that $T^*_{SR}(z, l^*b) \leq T^*_{Cusum}(z, l^*b)$.

$$E_\infty(\tilde{R}_n) = E_\infty\Big((1+\tilde{R}_{n-1})\frac{f_{l^*\lambda_1}(Y_n)}{f_{l^*\lambda_0}(Y_n)}\Big) = E_\infty(1+\tilde{R}_{n-1})E_\infty\Big[\frac{f_{l^*\lambda_1}(Y_n)}{f_{l^*\lambda_0}(Y_n)}\Big] = 1+E_\infty(\tilde{R}_{n-1}).$$

Thus, $E_\infty(\tilde{R}_n - n) = E_\infty(\tilde{R}_{n-1} - (n-1)) = \cdots = E_\infty(\tilde{R}_0 - 0) = e^z$. In fact, $\tilde{R}_n - n$ is a martingale because $E_\infty(M_n|M_{n-1}) = M_n$, in which $M_n = \tilde{R}_n - n$. Thus, $E_\infty(\tilde{R}_n - n) = e^z$ for any n.

By the martingale stopping theorem on page 300 of Ross (1996),

$$E_\infty[\tilde{R}_{T_{SR}(l^*b)}) - T_{SR}(l^*b)] = e^z,$$

which indicates that

$$E_\infty(T_{SR}(l^*b)) = E_\infty(\tilde{R}_{T_{SR}(l^*b)}) - e^z \geq e^{l^*b} - e^{l^*\eta b}.$$

Therefore, $E_\infty(T^*_{Cusum}(l^*b)) \geq E_\infty(T_{SR}(l^*b)) \geq e^{l^*b} - e^{l^*\eta b}$, and $\log E_\infty(T^*_{Cusum}(l^*b)) \geq \log(e^{l^*b} - e^{l^*\eta b}) = (1 + o(1))l^*b$. ■

Lemma **4.3** *As* $b \longrightarrow \infty$, $\log\Big(F(\eta b)P(T_{WLR}(b) \geq \omega)\Big) \longrightarrow 0$.

Proof.

$$\begin{aligned} F(\eta b)P(T_{WLR}(b) > \omega) &= P(0 < \tilde{W}_n < \eta b | T_{WLR}(b) > \omega)P(T_{WLR}(b) > \omega) \\ &= P(\tilde{W}_n < \eta b \text{ and } T_{WLR}(b) > \omega) \\ &\geq P_\infty(T_{Cusum}(l^{(0)}\eta b) > \omega) \end{aligned}$$

Thus,

$$\begin{aligned} 0 &\geq \log P_\infty\Big(T_{Cusum}(l^{(0)}\eta b) > \omega\Big) \\ &= -\frac{\omega}{e^{l^{(0)}\eta b}}. \end{aligned}$$

Because $l^*b + o(1) \geq \log \gamma$ in (45),

$$\log \left[F(\eta b) P\Big(T_{WLR}(b) > \omega \Big) \right] \quad \geq \quad -\frac{\omega}{\gamma^{\eta(\frac{l^{(0)}}{l^*})}} \to 0$$

by the assumption in Theorem 4.5. ■

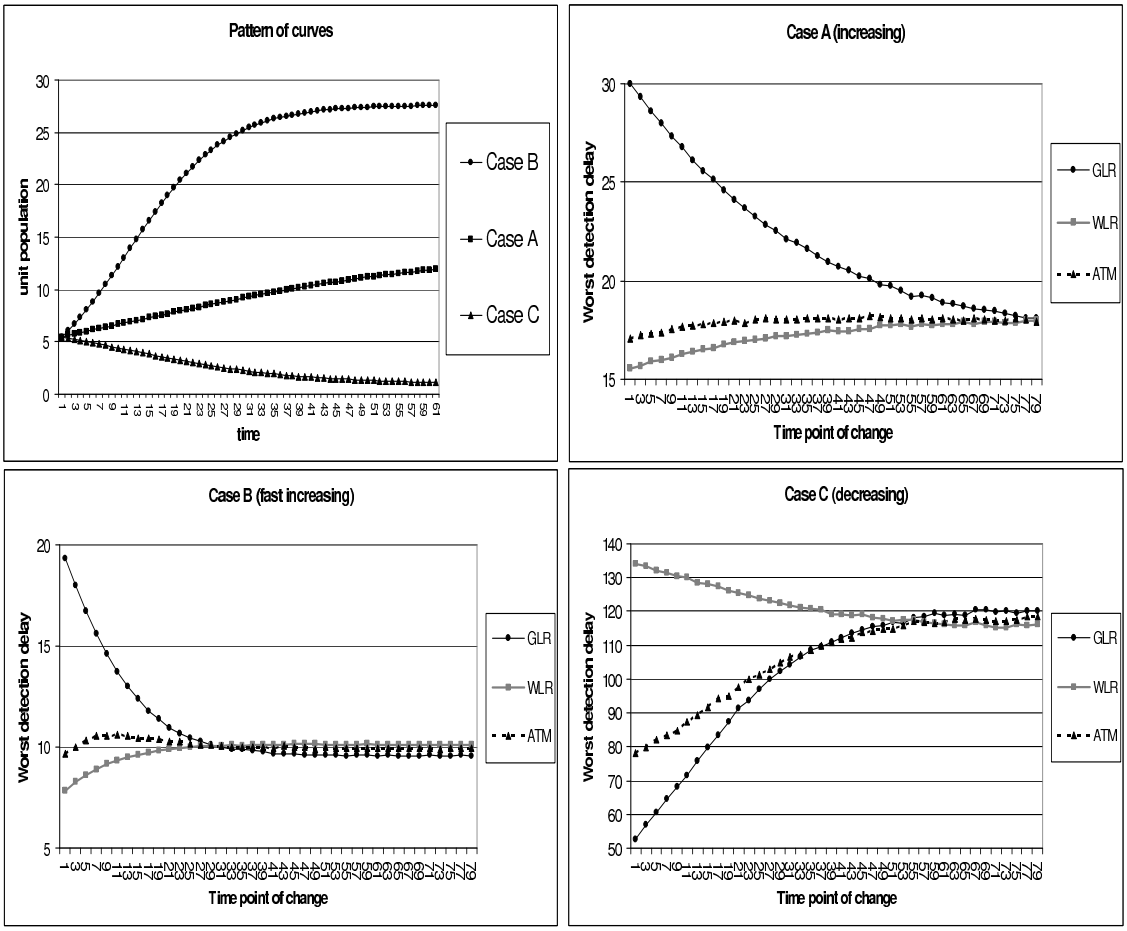


Figure 10: The population sizes are from the smooth model in (29). The *top left* panel plots three different population size curves that correspond to the three cases considered. The other three panels illustrate the detection delays of the three proposed detection schemes as a function of change-point ν under different cases of the population size models. The plots suggest that the WLR scheme $T_{WLR}(b)$ seems to be the best if the population sizes are increasing (Cases A and B), whereas the GLR scheme $T_{GLR}(a)$ seems to be the best if the population sizes are decreasing (Case C). Moreover, the ATM scheme $T_{ATM}(c)$ seems to be *robust* under Lorden's worst-case detection delay criterion no matter whether the population sizes are increasing or decreasing.

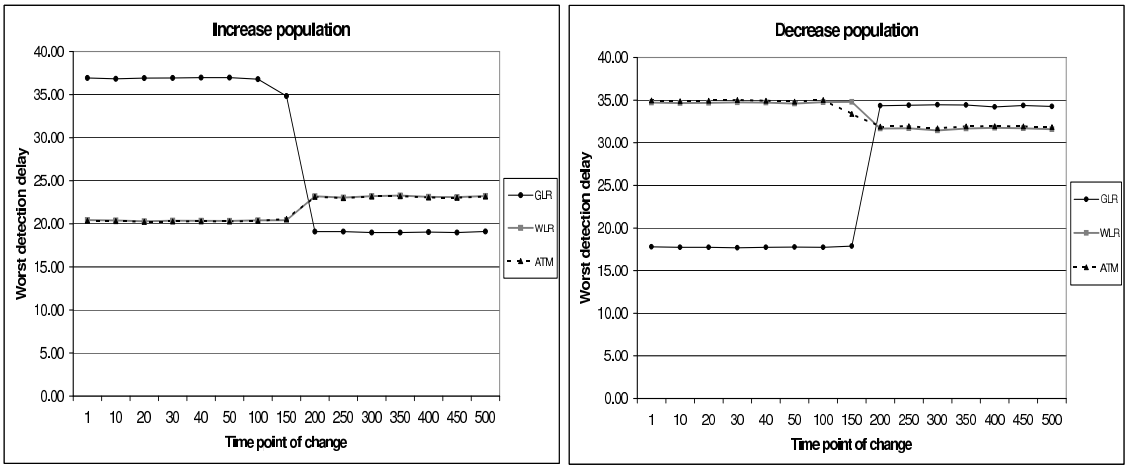


Figure 11: The detection delays of the three proposed detection schemes at different change-points ν when the population sizes are given by the step functions. Left Panel: the step function is increasing. Right Panel: the step function is decreasing.

CHAPTER V

SPATIOTEMPORAL SURVEILLANCE BASED ON LIKELIHOOD RATIO

Similar to the temporal CUSUM or scan statistic, the detection methods can be applied to spatial dimension. Raubertas [78] proposed the nearest neighborhood scan window, a method of scanning one region and closest neighbors of the region with fixed windows. Later, a variable scan-based method was developed by Kulldorff [51]. He suggested variable diameters of windows given each center and took the maximum value of log likelihood ratios from all possible areas. He did not assume incidence rates under both in-control and out-of-control states, so the log likelihood ratios are undefined. The nominator and denominator in the ratio are obtained by maximum likelihood. Later, Kulldorff [52] extended the spatial method to a spatiotemporal method by simply considering variable time periods.

Sonesson [94] invented CUSUM of circular scan statistics, which combines the ideas of Raubertas [78] and Kulldorff [52]. This method is used to detect emerging spatial clusters of disease. He defined temporal CUSUM statistics over all possible subsets of regions and took the maximum value of all temporal CUSUM statistics. He compared Kulldorff's scan statistics and CUSUM of circular scan with known parameters and concluded that the CUSUM procedure was superior. This conclusion is obvious because CUSUM is the optimal method under known baseline and change rates. In addition, Sonesson [94] compared his method (circular-CUSUM) with Raubertas's method (5 nearest neighborhood-CUSUM). The CED of circular-CUSUM is smaller than that of Raubertas, which indicates that the detection speed of circular-CUSUM is higher. Sonesson [94] and Frisén [26] discussed a number of these methods.

This chapter discusses spatiotemporal surveillance methods in a wide range of applications from industrial quality control to health surveillance and categorizes them to a general framework of likelihood ratios. We will compare their performance by the simulation study. The following sections of this chapter is the part of Tsui, Han, Jiang, and Woodall [103], which is submitted to *Technometrics*.

5.1 Introduction

Although many different surveillance methods (often referred to as control charts in industrial control) have been developed, we will focus our discussion on likelihood ratio (LR)-based methods, which are the building blocks of the cumulative sum (CUSUM) chart (Page [70]), popularly used in various applications. By investigating the CUSUM and other LR-based approaches in temporal surveillance for industrial quality control, we develop a framework consisting of many options for spatial and spatiotemporal surveillance. For simplicity, we assume that the sequential observations of the incidence rate are independent. As many spatiotemporal surveillance problems assume that the regional data of incidence rates are independent in both temporal and spatial aspects, we also assume the independence of observations across regions.

The chapter is organized as follows. Section 5.2 presents the notation for positions, coverage, and statistics. Section 5.3 discusses eight alternative surveillance methods for a spatiotemporal case under the independence assumption. Section 5.4 presents a real example based on the male thyroid cancer data in New Mexico during 1973-2006. Section 5.5 of this chapter contains our conclusions.

5.2 Notation for Positions, Coverage, and Statistics

For the notation of the spatial feature, we assume that adjacent regional data are independent. As shown in Figure 13, we define a spatial location by $c = (m, n)$, where $m = 1, 2, ..., M$, $n = 1, 2, ..., N$, and (m, n) represents the position coordinates in the

space dimension. Denote $\boldsymbol{R} = \{(m,n)\}$ all locations under consideration and $Y_c t$ the observation at location c at time t.

In order to simplify the problem, we assume that the outbreak coverage/cluster is a circle. We define an outbreak position by $c = (i,j)$ and outbreak coverage at a given position by A_c^r, where $A_c^r = \left\{ p = (m,n) \middle| \|p - c\| \leq r \right\}$ and r is the radius indicating the outbreak coverage or the cluster size.

As this chapter proposes different surveillance methods for spatiotemporal surveillance based on the likelihood ratio statistic, we will use notation T_{crt} as a testing statistic to represent LR statistics with different mathematical operations for certain possible sets of occurrence. Here the subscripts c, r, and t take values that represent certain operations over the center, radius, and shift times of the outbreak. Any of these subscripts can take value "M"(representing the "Max" operation), "S" (representing the "Sum" operation), "$-$" (representing no operation needed due to given knowledge), or an integer n (representing the size). When it takes "$-$", it represents a specific surveillance problem given certain knowledge. For example, T_{--M} represents the CUSUM procedure , and T_{--S} is the Shiryaev-Roberts procedure in the temporal surveillance case, where both the center and the radius of the outbreak is assumed to be known and the operation is taken over the LR statistics of all possible subsets of shift time; T_{M0M} and T_{M1M} are the Tartakovsky [100] and Raubertas [78] methods, respectively, in the spatiotemporal surveillance. When the shift magnitude is unknown and the generalized likelihood ratio statistics are used, then G instead of T is used as the general statistic. The specifics of these notation and surveillance methods will be discussed later in detail.

5.3 Spatiotemporal Surveillance Based on LR Statistics

In this section, we will discuss spatiotemporal surveillance methods based on the LR statistics of all possible windows. We first discuss the general sequential change detection problem for spatiotemporal surveillance. Then, we will propose a framework for spatiotemporal

scan statistics followed by their performance analysis.

5.3.1 Problem Formulation

In spatiotemporal surveillance, we assume that observations Y_{ct} are taken at location c at time t, where $c \in \boldsymbol{R}$, an area segmented into $M \times N$ regions. Assume that $Y_{ct} \sim f(\cdot|\mu_{ct})$, where μ_{ct} is the mean of Y_{ct}. When no outbreak occurs, it is assumed that $\mu_{ct} = \mu_0$ for all $c \in \boldsymbol{R}$, while $\mu_{ct} = \mu_1$ for some c when an outbreak occurs. Therefore, the spatiotemporal surveillance problem can be formulated as testing the null hypothesis

$$H_0 : \mu_{ct} = \mu_0 \text{ for all } c \text{ and } t \text{ (i.e., no change)}$$

against the composite alternative hypothesis

$$\begin{aligned} H_1 : \quad & \mu_{ct} = \mu_0 \text{ for all } c, \text{ when } t = 1, 2, ..., \nu - 1 \\ & \mu_{ct} = \mu_1 \text{ for some } c, \text{ when } t = \nu, \nu - 1, ..., \end{aligned}$$

where ν is an unknown time.

In order to test the above hypotheses, our framework for spatiotemporal surveillance is also based on the likelihood ratio statistic $\lambda_{ct}(\mu_0, \mu_1) = \frac{f_{\mu_1}(Y_{ct})}{f_{\mu_0}(Y_{ct})}$ for testing $\mu_{ct} = \mu_0$ against $\mu_{ct} = \mu_1$ at each location c and time t. For each position c, the windows for possible sets for likelihood ratios can be defined in two ways:

- Window of time: The cumulative likelihood ratio statistic over time $[\nu, t]$ is defined as $\Lambda^t_{c\nu} = \prod_{s=\nu}^{t} \lambda_{cs}(\mu_0, \mu_1)$.
- Window of radius: The cumulative likelihood ratio statistic over circle A^r_c is defined as $\Lambda^r_{ct} = \prod_{p \in A^r_c} \lambda_{pt}(\mu_0, \mu_1)$.

It is important to note that when the likelihood ratio statistics are aggregated over both windows of time and radius, the LR statistic is invariant to the order of time and radius operations, i.e.,

$$\Lambda^{rt}_{c\nu}(\mu_0, \mu_1) = \prod_{\nu=s}^{t} \Lambda^r_{c\nu} = \prod_{p \in A^r_c} \Lambda^t_{p\nu}.$$

5.3.2 Surveillance Methods Based on Likelihood Ratios

Similar to the pure temporal and spatial surveillance methods, we develop a set of scan statistics for spatiotemporal surveillance given different types of information regarding the change position c, coverage r, change magnitude μ_1, and most importantly, change time t.

5.3.2.1 When only the change position and the change time are unknown

Some spatiotemporal surveillance methods assume that change time ν and position c are unknown, but change coverage r and magnitude μ_1 are given. Tartakovsky [100] and Raubertas [78] considered using the maximum likelihood principle to estimate the unknown parameters and developed special cases of the following scan statistic

$$T_{MrM}(t) = \max_{c} \Big\{ \max_{1\le\nu\le t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0, \mu_1) \Big\} \Big\}, \tag{46}$$

where the first "M" in the subscript stands for the maximum over all regions $c \in \boldsymbol{R}$ and the second one stands for the maximum over all possible time windows $[\nu, t]$. Tartakovsky [100] considered a single region, i.e., r=0, while Raubertas [78] studied the same statistic assuming that the shift coverage is nearest neighborhood regions with $r = 1$.

By applying the summation operation instead of the maximum, we can develop the following three spatiotemporal scan statistics

$$T_{MrS}(t) = \max_{c} \Big\{ \sum_{\nu=1}^{t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0, \mu_1) \Big\} \Big\} \tag{47}$$

$$T_{SrM}(t) = \sum_{c} \Big\{ \max_{1\le\nu\le t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0, \mu_1) \Big\} \Big\} \tag{48}$$

$$T_{SrS}(t) = \sum_{c} \Big\{ \sum_{\nu=1}^{t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0, \mu_1) \Big\} \Big\}. \tag{49}$$

Mei [62] considered T_{S0M} and found that T_{S0M} is competitive with T_{M0M}.

It is important to note that exchanging the time and position in Eqs. (46) and (49) retains the same statistics. However, exchanging the time and position over the maximum and summation operations in Eqs. (47) and (48) may yield different statistics. For example,

for the $T_{SrM}(t)$ statistic,

$$\sum_c \Big\{ \max_{1\le\nu\le t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \neq \max_{1\le\nu\le t} \Big\{ \sum_c \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\}.$$

Closely looking at the statistics, we found that the right-hand side expression seems inconsistent with the change point assumption that at a certain time point an outbreak occurs in a certain region. Therefore, we suggest taking (max/sum) operations over time for fixed locations.

5.3.2.2 When change position, coverage, and time are unknown

Sonesson [94] assumed that the change position c, coverage r, and time ν are all unknown except when the magnitude μ_1 is given, and considered the maximum operation over all unknown parameters. By applying the maximum and summation operations to the three unknown parameters interchangably, we can obtain the following eight scan statistics

$$T_{MMM}(t) = \max_c \Big\{ \max_r \Big\{ \max_{1\le\nu\le t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \tag{50}$$

$$T_{MMS}(t) = \max_c \Big\{ \max_r \Big\{ \sum_{\nu=1}^{t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \tag{51}$$

$$T_{SMM}(t) = \sum_c \Big\{ \max_r \Big\{ \max_{1\le\nu\le t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \tag{52}$$

$$T_{SMS}(t) = \sum_c \Big\{ \max_r \Big\{ \sum_{\nu=1}^{t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \tag{53}$$

$$T_{MSM}(t) = \max_c \Big\{ \sum_r \Big\{ \max_{1\le\nu\le t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \tag{54}$$

$$T_{MSS}(t) = \max_c \Big\{ \sum_r \Big\{ \sum_{\nu=1}^{t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \tag{55}$$

$$T_{SSM}(t) = \sum_c \Big\{ \sum_r \Big\{ \max_{1\le\nu\le t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\} \tag{56}$$

$$T_{SSS}(t) = \sum_c \Big\{ \sum_r \Big\{ \sum_{\nu=1}^{t} \Big\{ \Lambda_{c\nu}^{rt}(\mu_0,\mu_1) \Big\} \Big\}, \tag{57}$$

where A_c^r can include mostly 50% of the whole area from the center. Note that Eq. (50) corresponds to Sonesson's method.

5.3.2.3 *When the change position, coverage, magnitude, and time are unknown*

Kulldorff [52] assumed that the change time, position, coverage, and magnitude are all unknown and suggested estimating the shift magnitude by the maximum likelihood principle. In general, the eight spatiotemporal statistics can be extended to handle the case when μ_1 is unknown by maximizing the aggregated LR statistic $\Lambda_{cv}^{rt}(\mu_0, \mu_1)$ with respect to μ_1.

5.3.3 Performance Comparison

We now compare the performance of the aforementioned scan statistics for spatiotemporal surveillance. We use the average run length (ARL), the average number of observations before a signal is triggered, as the criterion for comparison. We fix the in-control ARL–the ARL when no outbreak occurs–to 100 and compare the out-of-control ARL (also referred to as conditional expected delay (CED))–the ARL with outbreaks. The smallest out-of-control ARL indicates the best scan statistic.

Similar to spatial surveillance, the following comparison was conducted on a 6×6 regular map, shown in Figure 13, and the outbreak patterns in Figure 12 were simulated to see which scan statistic provides the shortest ARL. The simulation was repeated 100,000 times for in-control ARL_0 and 10,000 times for out-of-control ARL until the standard error of the ARL estimation was negligible. To compare the variations of performance measures, we ran the simulation in both zero- and steady-state. The zero-state assumes that outbreaks occur at the beginning of the initialization of the scan statistics while the steady-state refers to the situation when the scan statistics have been monitored for a sufficiently long period of time (50 observations in our simulation).

5.3.3.1 *When only the change position and time are unknown*

Assuming that only the change position is unknown, we use a different radius to compare the scan statistics in Eqs. (46) - (49). Table 6 shows both steady- and zero-state ARL for these scan statistics when the outbreak exhibits patterns of S-1, S-5, S-9, and S-13 with a

change magnitude of 0.5. The smallest ARL among the different statistics is highlighted in bold. Generally, the best performance can be attained when the scan radius matches the outbreak coverage, which is consistent with our expectations. However, neither Tartakovsky's method nor Raubertas's method is the best among these scan statistics in either the steady or the zero state. Replacing the maximum operation by the summation in the spatial and/or temporal windows can often improve the detection of outbreaks.

For the steady state when the scan statistics have been running for 50 periods, taking the summation in both spatial and temporal windows always results in better performance when the change coverage is small, regardless of whether the radius is over-scanned or under-scanned. The improvement can be 6-7% when r is small. On the other hand, when the change coverage is large, taking the summation over the spatial window but the maximum over the time window often generates the best ARL results. The improvement is more significant when the coverage is seriously under-scanned, which can be as large as more than 20%. Note that this scan statistic is identical to the one Mei suggested [62]. It not only offers superior performance for the steady state detection, but also gives the best performance.

Although we only present the case in which the change magnitude is 0.5, similar experiments for the change magnitude of 1 and 2 show similar comparisons, while the difference among these scan statistics becomes smaller when the magnitude is large. Nevertheless, when an outbreak massively spreads out, Mei's method is always better than Tartakovsky's method. When the radius r becomes larger, the difference between Eqs. (46) and (48) is minor.

5.3.3.2 When the change position, radius and time are all unknown

Table 7 shows the ARL values for the scan statistics in Eqs. (50)-(57) in both steady and zero states when a search of the radius is used. Due to random errors in the simulation, it is not easy to pinpoint the best monitoring statistic. However, it is easy to see that T_{SMM},

Table 6: ARL of the scan statistics when the change position and time are unknown (the change magnitude is 0.5)

Methods	S-1	S-5	S-9	S-13
		Steady-state		
T_{M0M}	27.25	14.92	12.19	10.74
T_{S0M}	26.75	13.25	10.01	8.46
T_{M0S}	25.90	14.66	12.01	10.66
T_{S0S}	**25.30**	**12.98**	**9.87**	**8.33**
T_{M1M}	54.21	8.68	6.49	5.27
T_{S1M}	53.39	8.54	**6.00**	**4.72**
T_{M1S}	52.13	8.75	6.62	5.37
T_{S1S}	**50.89**	**8.54**	6.08	4.84
T_{M2M}	63.32	11.84	5.14	4.44
T_{S2M}	63.21	11.68	**5.06**	**4.14**
T_{M2S}	63.14	11.70	5.19	4.52
T_{S2S}	**62.09**	**11.41**	5.15	4.23
		Zero-state		
T_{M0M}	35.27	19.96	17.05	15.61
T_{S0M}	**35.22**	**18.76**	**15.17**	**13.31**
T_{M0S}	38.45	23.56	20.72	19.27
T_{S0S}	38.63	22.60	19.10	17.16
T_{M1M}	57.86	9.76	7.57	6.19
T_{S1M}	57.11	**9.69**	**7.11**	**5.69**
T_{M1S}	57.64	10.43	8.31	6.96
T_{S1S}	**56.83**	10.41	7.91	6.50
T_{M2M}	66.41	12.83	5.58	4.93
T_{S2M}	**65.61**	**12.73**	**5.52**	**4.69**
T_{M2S}	67.36	13.29	6.00	5.34
T_{S2S}	67.03	13.08	5.97	5.11

Table 7: ARL of the scan statistics when the change position, radius, and time are all unknown (the change magnitude is known to be 0.5)

Methods	S-1	S-5	S-9	S-13
		Steady-state		
T_{MMM}	33.55	9.11	5.46	4.13
T_{SMM}	35.23	9.26	5.32	3.85
T_{MMS}	29.09	9.37	5.84	4.47
T_{SMS}	30.42	9.10	5.54	4.07
T_{MSM}	34.72	9.32	5.51	4.12
T_{SSM}	36.74	9.52	5.42	3.88
T_{MSS}	30.39	9.39	5.84	4.40
T_{SSS}	31.89	9.23	5.44	4.01
		Zero-state		
T_{MMM}	37.46	10.10	6.02	4.47
T_{SMM}	38.79	10.14	5.86	4.31
T_{MMS}	37.74	11.13	6.78	5.11
T_{SMS}	38.04	11.07	6.56	4.89
T_{MSM}	38.58	10.29	6.02	4.52
T_{SSM}	40.48	10.47	5.90	4.36
T_{MSS}	38.24	11.19	6.72	5.09
T_{SSS}	38.97	11.23	6.54	4.89

which is generalized from T_{SM-} from spatial to spatiotemporal surveillance, provides the best ARL performance for outbreaks with a large coverage while T_{SMS} provides nearly the best performance for outbreaks with a small coverage. Nevertheless, the zero-state ARL shows that Sonesson's method works the best for outbreaks with a small coverage.

5.4 New Mexico Thyroid Cancer Example

The following example is based on male thyroid cancer data in New Mexico from 1973 to 2006 based on Figure 1, Section 2.4. We investigated male thyroid cancer incidence with malignant behavior in 32 counties in New Mexico. The map of the counties in New Mexico is in the left panel of Figure 15, which is referred to on the web site of the U.S. Census Bureau (http://quickfacts.census.gov/qfd/index.html).

In this numerical example, as Kulldorff and Sonesson did, we assumed that the observations follow a Poisson distribution defined by

$$f_{\pi_{ct}}(y_t) = \frac{\pi_{ct}^{y_{ct}} e^{-\pi_{ct}}}{y_{ct}},$$

where $\pi_{ct} = N_{ct}\mu_{ct}$ is the incidence rate of country c at time t, N_{ct} and μ_{ct} are the corresponding population size and incidence rate per 100,000 males. Our objective is to detect as quickly as possible an increased rate, e.g., 50% increase from the baseline in any region. Thus, we formulated the problem as follows,

$$H_0 : \mu_{ct} = \mu_0 \text{ for all regions } c \text{ and time } t \text{ (i.e., no change)}$$

against

$$\begin{aligned} H_1 : \quad & \mu_{ct} = \mu_0 \text{ for all regions } c, \text{ when } t = 1, 2, ..., \nu_0 - 1 \\ & \mu_{ct} = \mu_0(1 + 0.5) \text{ for some regions } c, \text{ when } t = \nu_0, \nu_0 - 1, ..., \end{aligned}$$

where ν_0 is an unknown time.

We first estimated a baseline rate for each region. Figure 16 shows the yearly incidence of male thyroid cancer per 100,000 males in New Mexico. It is easy to see that the yearly incidence rate gradually increased since 1989, especially the monotone upward trend after 1995, while no obvious change occurred from 1973 to 1988. Therefore, we used the data between 1973 to 1988 to estimate the baseline rate in each county, $\mu_0 = 2.4$. We noticed that the difference in the population among the counties was large; thus we assume that the baselines for different counties are non-homogeneous, and the total incidence rate of county c at time t under normal state is $N_{ct}\mu_{ct}$.

After estimating the baseline rates, we attempted to understand which counties had increased incidence rates from 1989 to 2001. We compared the average incidence rates of 1973-1988 with those of 1989-2001 and shaded the areas that had 50% increases, as shown in Figure 15 (a). In order to quickly detect the responsible clusters whose incidence had increased 50% in an online fashion, we applied the spatiotemporal surveillance methods

discussed in Section 5.3 to monitor the changes in incidence rates. Due to the irregular patterns in the map, we had to adjust the cluster definition and position the location of centers in counties to X-Y coordinates as shown in Figure 15 (b). The cluster around each center was defined as those centers that had a Euclidean distance of less than radius r in the X-Y coordinates. In this study, we considered (i) $r = 5$ and (ii) variable r to cover at most 50% of the entire population in the scan statistics.

To fairly compare the spatiotemporal scan statistics, we performed a Monte Carlo simulation study to compare the ARL values of different scan statistics. We set $ARL_0 = 100$ and determined the threshold using 100,000 replications. For $r = 5$, the control limits for T_{M5M}, T_{S5M}, T_{M5S}, and T_{S5S} were 162, 403.4, 891, and 2162, respectively. Since the scan statistics had different scales and control limits, Figure 17 plots the scan statistics from 1989 to 2006 scaled by their corresponding control limits. A signal was triggered when a scaled statistic exceeded 1. In the comparison of the different scan statistics, all methods signaled around year 1994-1995 when r was considered a variable. When r was assumed to be 5, most methods signaled from 1993 to 1995. In the $r = 5$ case, T_{S5M} detected the increased rate most quickly, and in the variable r case, T_{MMM}, T_{SMM}, and T_{SSM} detected the increase faster than the other. These results also indicate that the maximum operation for the temporal domain and the summation operation for the location part are the best combination.

In addition to the conditional delay, we investigated the clusters that most strongly contributed to the alarm. For $r = 5$, Figure 18 depicts the clusters, including Bernalillo, Sandoval, and Los Alamos Counties, that were identified by the surveillance methods. This observation seems consistent with what Kulldorff (2001) observed as the emerging cluster around Los Alamos. In fact, a single region scan test (r=0) may easily signal Los Alamos county as Kulldorff did. Furthermore, for a variable radius, Figure 19) shows that more counties were responsible for the increased incidence around Bernalillo, Sandoval, and Los Alamos Counties.

Although minor differences occur among different surveillance methods, as discussed in Sonesson (2007), it is hard to to determine whether the alarm results from a true clustering or by chance and epidemiologists need to investigate clusters in practice. More study is needed to further compare the performance of different scan statistics.

5.5 Conclusion

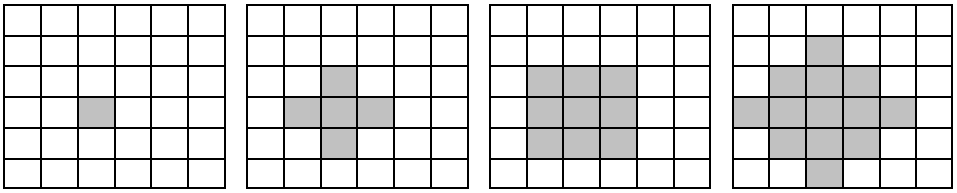
In this chapter, we developed a LR-based framework for spatiotemporal surveillance that included many existing methods by taking different operations over the LR statistics in variable windows. We compared the power and performance of these scan statistics for both spatiotemporal surveillance under different assumptions on the change position, coverage, and magnitude. A few key findings from this research are summarized as follows:

1. Similar to the Shiryaev-Roberts approach in temporal surveillance, taking the summation for spatiotemporal surveillance may outperform the conventional procedure, which takes the maximum over all windows of the LR statistics.

2. When the outbreak coverage is known, scan statistics with an appropriate radius that matches the actual outbreak coverage often perform better than the statistics under- or over-scanned.

3. When searching over various radii in the scan statistics, taking the summation over spatial windows but the maximum over temporal windows generally provides better performance than other approaches.

Although normal distributions are assumed in our simulation study, the general LR statistics defined in the framework are valid for other distributions such as Poisson in the real example. While we assume independent observations in each location both spatially and temporally, more extensive work is needed towards more general assumptions including cross-sectional data or even panel data, both of which are common in econometric research [4]. This will be further pursued in future research.



(a) S-1 (b) S-5 (c) S-9 (d) S-13

Figure 12: Outbreak patterns in the simulation

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Figure 13: Vector and coordinate expressions of regional data (p=36)

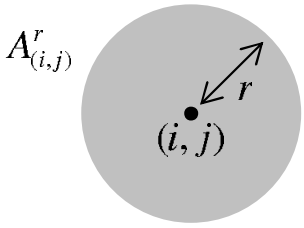
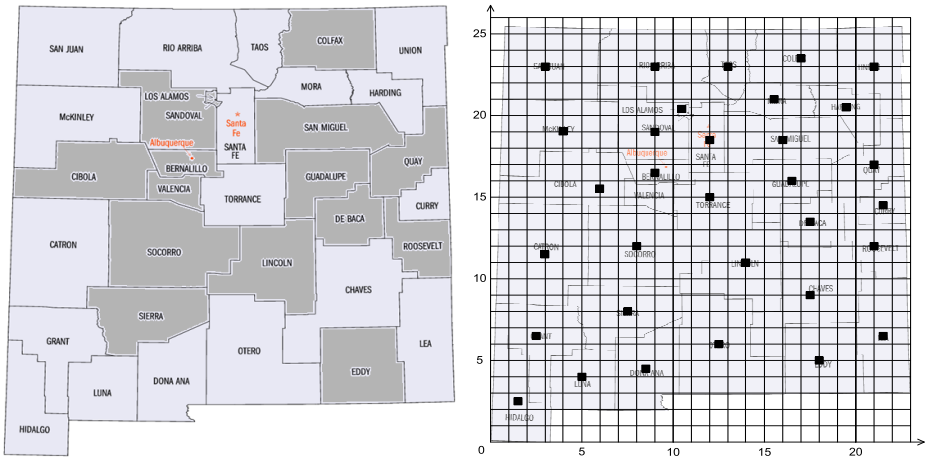


Figure 14: Outbreak position $c = (i, j)$, radius r, and coverage A^r_c



(a) The map of counties in New Mexico (b) X-Y coordinate representation

Figure 15: The map of counties in New Mexico and the X-Y coordinate representation

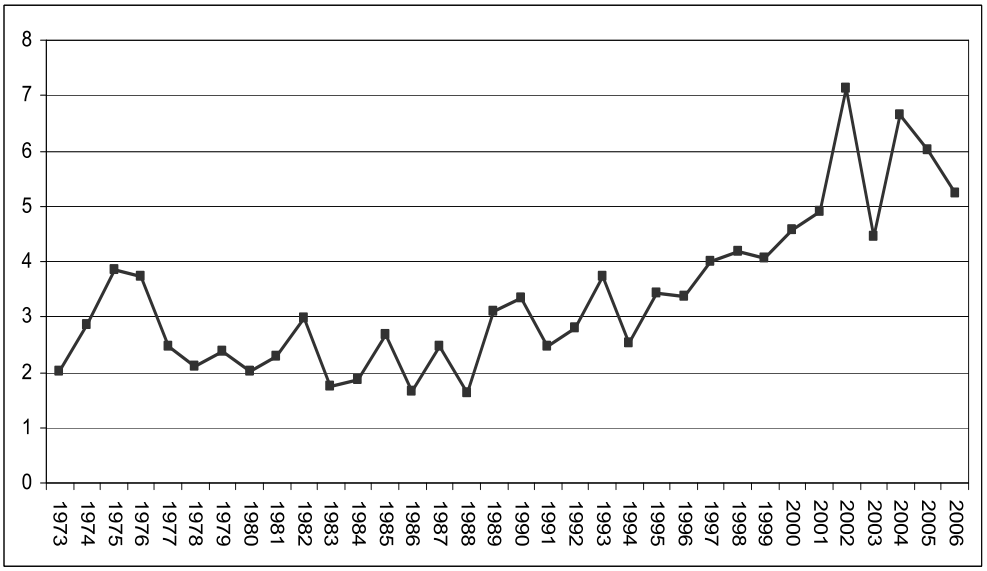
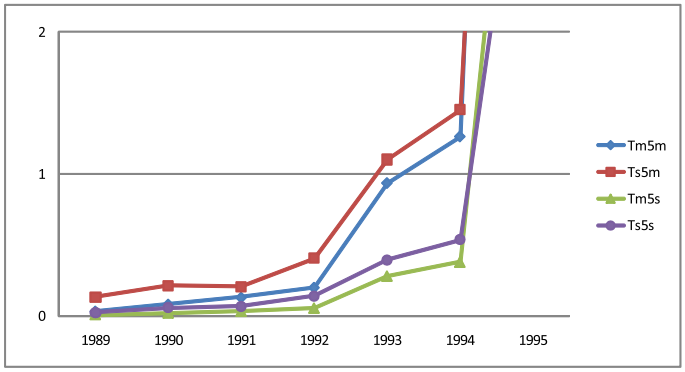
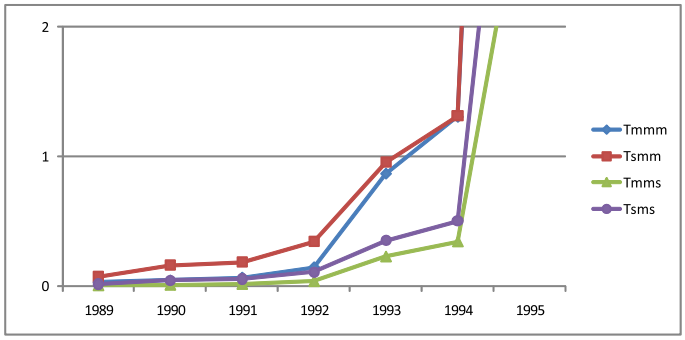


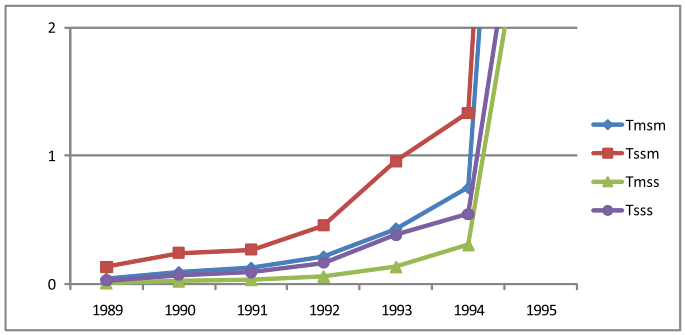
Figure 16: Male thyroid cancer incidence per 100,000 in New Mexico



(a) $r = 5$

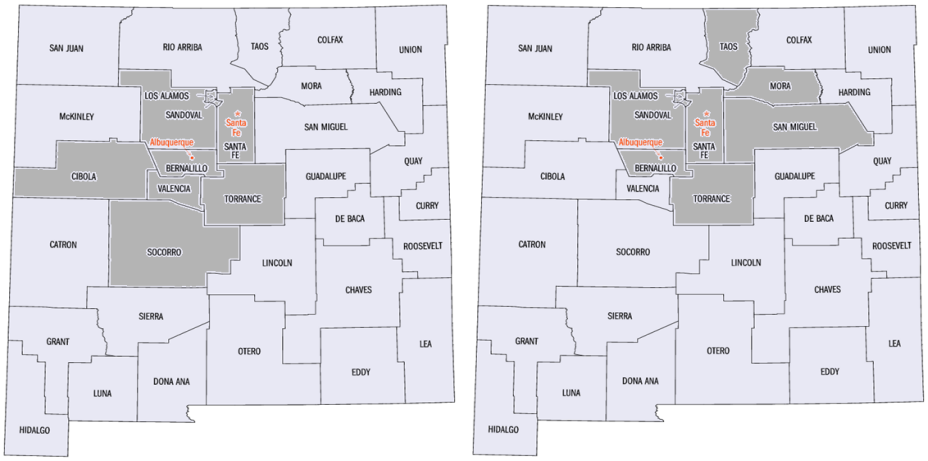


(b) Variable r with maximum over different radii



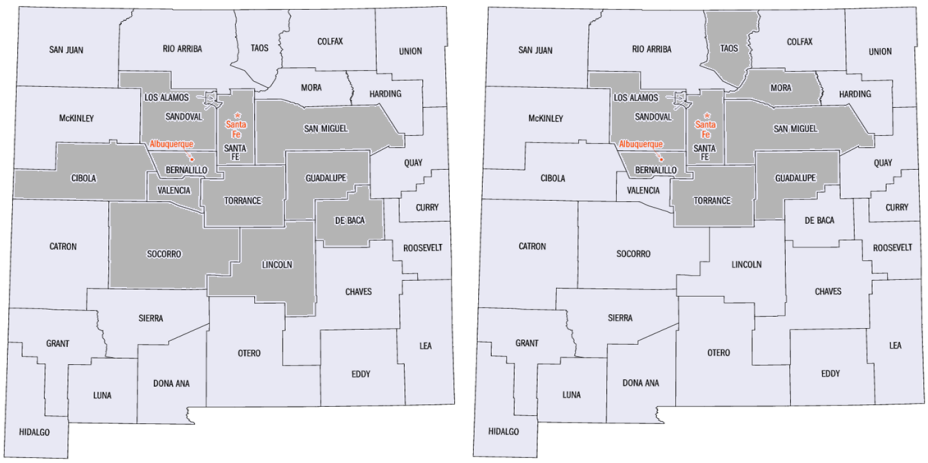
(c) Variable r with summation over different radii

Figure 17: Scaled scan statistics for male thyroid cancer data in New Mexico



(a) T_{-5M} (b) T_{-5S}

Figure 18: Clusters detected for the male thyroid cancer data in r=5 cases



(a) T_{--M} except for T_{-SS} (b) T_{-SS}

Figure 19: Clusters detected for the male thyroid cancer data in variable radius cases

CHAPTER VI

SPATIOTEMPORAL SURVEILLANCE METHODS IN THE PRESENCE OF SPATIAL CORRELATION

For multivariate surveillance with correlated data, an underline distribution has been defined as a multinormal distribution that specifies the mean vector and the covariance matrix. Under the distribution assumption, either CUSUM or EWMA is used, and likelihood ratios or the Mahalanonobis distance (Hotelling T^2) is selected as a distance measure.

For CUSUM procedures, Woodall *et al.* [112] considered multiple temporal CUSUM statistics and took the maximum overall statistics. Crosier [16] and Pignatiello *et al.* [72] suggested two methods. In the first method, the T^2 statistic is constructed from the observation vector, and the CUSUM of T^2 is formed. In the second method, multivariate CUSUM statistics are contructed from corresponding observations, and the T^2 of the CUSUM variable is calculated. The two methods in Crosier [16] are slightly different with respect to the technique from those in Pignatiello *et al.* [72]. Healy [36] used log likelihood ratios of a multivariate normal distribution to construct the CUSUM statistic. Hawkins [33, 34] decomposed the correlated observations into independent residuals by linear regression, called the regression-adjusted method.

Multivariate surveillance based on SPC was adapted to the spatiotemporal context by Rogerson [85] and Rogerson *et al.* [87]. Rogerson [85] introduced control chart-based work on time and regions. He applied the Tango [99] to an on-line monitoring application using a CUSUM procedure. Rogerson [86] modified the retrospective cluster detection method of Knox [50] to the prospective case. The statistical performance of this method was evaluated by Marshall *et al.* [61], who showed that simulation was needed to determine its in-control performance and to design the monitoring scheme. Rogerson *et al.* [87] and

[88] considered the spatiotemporal surveillance problem in which the count data in the regions are correlated. They compared the performance of multiple CUSUM charts (Woodall *et al.* [112]) with that of the multivariate CUSUM procedure (Crosier [16]). Testik *et al.* [101] provided some very recent work on one-sided multivariate control charts. Kleinman *et al.* [49] and Kleinman [48] also considered the spatiotemporal aggregated case, they recommended predicting the incidence rate for each sub-region and using recurrence intervals to assess whether an alarm is given or not.

Finally, Jiang and Tsui [44] related the T^2 chart, the M-chart, and the regression-adjusted chart and showed that the performance of the methods depends on the direction and the size of the shifts as well as the correlation magnitudes. They introduced a hybrid method that combines the T^2 chart and the regression-adjusted chart to improve performance.

For EWMA procedures, Lowry *et al.* [58] presented a multivariate extension of the EWMA procedure by constructing the T^2 statistic over the EWMA statistics. Tsui *et al.* [104] designed a multivariate EWMA control chart based on the loss functions from the vector of responses. Joner *et al.* [47] proposed a one-sided multivariate EWMA as a charting T^2 statistic over the one-sided EWMA statistic. Similar to the research in multivariate normal distribution, an investigation of various multivariate CUSUM and EWMA procedures under multivariate Bernoulli and Poisson data is of interest for health surveillance.

In this chapter, we discuss spatiotemporal surveillance methods in the presence of spatial correlations. The following sections of this chapter is an excerpt from Jiang, Han, Tsui, and Woodall [43], which will be published in *Statistics in Medicine*, DOI: 10.1002/sim.3877. The excerpt format is based on the Copyright Transfer Agreement (2009) in Wiley Blackwell, and some grammar corrections have been made.

6.1 Introduction

Public health data often consist of spatial information on the location of incidences. As many researchers pointed out [52, 94], an increase in incidence often occurs in a cluster of regions. Many spatiotemporal surveillance methods are based on the assumption that the regional observations are independent; however, they are, in fact, often correlated due to similarities among adjacent regions. Monitoring correlated spatial observations has not been thoroughly studied by the public health surveillance research community.

In industrial process control, correlated multivariate observations are frequently monitored to detect and identify critical process characteristics or processes that are responsible for shifts of product quality. The underlying distribution has often been assumed to be a multivariate normal distribution with a specified mean vector and covariance matrix. Under the multivariate normal distribution assumption, Jiang and Tsui [44] reviewed Shewhart-type multivariate process control charts in a unified framework of likelihood ratio tests. To improve the detection of small sustained mean changes, multivariate CUSUM (MCUSUM) or EWMA (MEWMA) charts have been proposed in the statistical process control (SPC) community. In this chapter, we will discuss various multivariate CUSUM-type methods for spatiotemporal surveillance and compare their performance under different scenarios. By taking advantage of the characteristics of an outbreak cluster, we propose several alternative CUSUM-type methods for spatiotemporal surveillance.

The contents of this chapter are as follows. Section 6.2 develops a likelihood-ratio-based approach for spatiotemporal surveillance for correlated data, and Section 6.3 compares the performance of the detection methods. Section 6.4 presents an example of breast cancer in New Hampshire to illustrate the effectiveness of different surveillance methods when spatial correlations are present. Finally, Section 6.5 summarizes the findings and discusses some future research. The appendix provides some technical derivations of the proposed methods based on the likelihood ratio principle.

6.2 MCUSUM Methods for Spatiotemporal Surveillance

6.2.1 Problem Definition

When observations from each spatial region are taken as components of a multivariate vector, spatiotemporal surveillance is similar to multivariate surveillance, except that additional information about change coverage or clusters has to be examined [103]. As shown in Figure 13, we define a spatial location by $c = (m, n)$, where $m = 1, 2, ..., M$, $n = 1, 2, ..., N$, and (m, n) represents the position coordinates in the space dimension. We denote all locations under consideration by $\boldsymbol{R} = \{(m, n)\}$ and the observation from location c by x_c. We assume that observations x_{ct} are taken from location c at time t, where $c \in \boldsymbol{R}$, an area that is segmented into $p = M \times N$ regions. We assume that $x_{ct} \sim f(\cdot|\mu_{ct})$, where μ_{ct} is the mean of x_{ct}.

In most health surveillance research, it is customary to assume that $\mu_{ct} = \mu_0$ for all $c \in \boldsymbol{R}$ when no outbreak occurs, while $\mu_{ct} = \mu_1$ (constant with respect to c) for some regions when an outbreak occurs. Therefore, the spatiotemporal surveillance problem can be formulated as testing the null hypothesis

$$H_0 : \mu_{ct} = \mu_0 \text{ for all } c \text{ and } t \text{ (i.e., no change)}$$

against the composite alternative hypothesis

$$\begin{aligned} H_1 : \quad & \mu_{ct} = \mu_0 \text{ for all } c, \text{ when } t = 1, 2, ..., \nu - 1 \\ & \mu_{ct} = \mu_1 \text{ for some } c, \text{ when } t = \nu, \nu - 1, ..., \end{aligned}$$

where ν is an unknown time. We refer to cases where $\mu_{ct} = \mu_1$ as homogeneous outbreaks because all regions in a cluster have an identical mean shift.

In order to simplify the problem, we assume that the outbreak coverage/cluster is a circle. We define an outbreak center by $c = (i, j)$ and outbreak coverage around c by A_c^r, where $A_c^r = \left\{ k = (m, n) \middle| \|k - c\| \leq r \right\}$ and r is the radius indicating the outbreak coverage or the cluster size. Here k represents any region with a midpoint covered by a circle with radius r from center c. See Figure 14 for an illustration.

To test the above clustering hypotheses, our framework for spatiotemporal surveillance is also based on the likelihood ratio statistic on any cluster A_c^r from time ν to t. Let $\mathcal{P} = \{1, 2, \ldots, p\}$ be the set of all variable indices. Denote $\boldsymbol{\Delta}_{A_c^r}$ the vector space with nonzero values at and only at the index on subset A_c^r and $(A_c^r) = \mathcal{P} \backslash A_c^r$ the complementary subset of A_c^r. In health surveillance applications, the nonzero component of vector $\boldsymbol{\Delta}_{A_c^r}$ is $\delta = \mu_1 - \mu_0$. In the following, we will apply both UIT and GLRT principles (Jiang and Tsui [44]) to develop monitoring statistics over all regions for spatiotemporal surveillance given different types of information regarding the change position c, coverage r, and change time ν.

6.2.2 When the Change Position and Time Are Unknown

As discussed before, some spatiotemporal surveillance methods are based on the assumption that change time ν and position c are unknown, but change coverage r and magnitude μ_1 are given. Under the independence assumption, Tartakovsky [100] proposed monitoring the cumulative sum of the LR statistics for each region, which is equivalent in form to Woodall and Ncube's MCUSUM statistic [112]. Rogerson and Yamada [88] also considered this single-region-based MCUSUM scheme when correlations are present. However, both papers used the UIT principle, but they did not take into account the correlations among regions. In general, when testing the significance of cluster A_c^r with radius r around region c, the likelihood ratios based on the GLRT principle is reduced to

$$T_t(r) = \max_c \max_{1 \le \nu \le t} L_\nu^t(A_c^r), \tag{58}$$

where $L_\nu^t(A_c^r) = \delta_{A_c^r}\left[\boldsymbol{\Delta}'_{A_c^r}\boldsymbol{\Sigma}^{-1}\boldsymbol{s}_\nu^t/||\boldsymbol{\Delta}_{A_c^r}||_{\boldsymbol{\Sigma}} - \delta_{A_c^r}(t-\nu+1)/2\right]$ and $\delta_{A_c^r} = ||\boldsymbol{\Delta}_{A_c^r}||_{\boldsymbol{\Sigma}}$. Note that the two maximum operations can be exchanged; however, recursive forms exist for each region c only, which is easy to compute.

6.2.2.1 When homogeneous changes are considered

For example, as we mentioned earlier, it is common to assume $\boldsymbol{\Delta}_{A_c^r} = \delta \mathbf{1}_{A_c^r}$ in health surveillance applications, where δ is given and $\mathbf{1}_{A_c^r}$ represents a p-dimensional vector with 1's only in the index position of the regions inside A_c^r but 0 elsewhere, i.e., the shift magnitudes are assumed to be the same across all regions in A_c^r. We let $\boldsymbol{D} = \left(\boldsymbol{\Sigma}_{A_c^r} - \boldsymbol{\Sigma}_{A_c^r(A_c^r)}\boldsymbol{\Sigma}^{-1}_{(A_c^r)(A_c^r)}\boldsymbol{\Sigma}_{(A_c^r)A_c^r}\right)^{-1}$ and $\boldsymbol{F} = \boldsymbol{D}\boldsymbol{\Sigma}_{A_c^r(A_c^r)}\boldsymbol{\Sigma}^{-1}_{(A_c^r)(A_c^r)}$. The LR statistic in Eq. (58) reduces to

$$L_\nu^t(A_c^r) = \delta(A_c^r)\big[u(A_c^r) - \frac{\delta(A_c^r)}{2}(t-\nu+1)\big], \tag{59}$$

where $u(A_c^r) = \mathbf{1}'\boldsymbol{D}\boldsymbol{z}(A_c^r)/\sqrt{\mathbf{1}'\boldsymbol{D}\mathbf{1}}$, $\delta_{A_c^r} = \delta||\mathbf{1}_{A_c^r}||_{\boldsymbol{\Sigma}}$, and

$$\boldsymbol{z}(A_c^r) = (\boldsymbol{s}_\nu^t)_{A_c^r} - \boldsymbol{\Sigma}_{A_c^r(A_c^r)}\boldsymbol{\Sigma}^{-1}_{(A_c^r)(A_c^r)}(\boldsymbol{s}_\nu^t)_{(A_c^r)} = (\boldsymbol{s}_\nu^t)_{A_c^r} - \boldsymbol{D}^{-1}\boldsymbol{F}\cdot(\boldsymbol{s}_\nu^t)_{(A_c^r)},$$

which is the regression adjustment of cluster $(\boldsymbol{s}_\nu^t)_{A_c^r}$ with respect to the rest $(\boldsymbol{s}_\nu^t)_{(A_c^r)}$ and has variance-covariance matrix $\boldsymbol{D}^{-1}$. When $r = 0$, the above statistic reduces to Hawkins' regression-adjusted method for individual regions. In this chapter, we will consider more general radius $r > 0$, which represents regression-adjusted clusters.

6.2.2.2 When the coverage is unknown

Under the independence assumption, Sonesson [94] derived a monitoring statistic, assuming that the change time, position, and coverage are unknown except for shift magnitude δ. By maximizing the statistics over radius r, we can obtain a monitoring statistic based on the generalized LR test with spatial correlations as follows

$$T^S = \max_c \max_r \max_{1\le\nu\le t} L_\nu^t(A_c^r), \tag{60}$$

where $L_\nu^t(A_c^r)$ follows from Eq. (59). Radius r is chosen so that A_c^r can include at most 50% of the entire area from c.

6.2.3 When the Change Position, Coverage and Magnitude Are Unknown

More generally, under the independence assumption, Kulldorff's spatial statistic [52] allows the change time, position, coverage, and magnitude to all be unknown. He suggested estimating the change magnitude using the maximum likelihood principle. For each cluster A_c^r, assuming the change occurs at time ν, $\boldsymbol{\Delta}_{A_c^r}$ in Eq. (58) can be estimated as $\hat{\boldsymbol{\Delta}}_{A_c^r} = (\boldsymbol{z}'(A_c^r)/(t-\nu+1), \mathbf{0}'_{(A_c^r)})'$, i.e., the regression-adjusted cluster of A_c^r. In this case, Eq. (58) is reduced to

$$L_\nu^t(A_c^r) = \frac{||\boldsymbol{z}(A_c^r)||^2_{\boldsymbol{D}^{-1}}}{2(t-\nu+1)}. \tag{61}$$

Again, if homogeneous changes are considered across cluster regions but the change magnitude δ is unknown, δ can be estimated as $\hat{\delta} = u(A_c^r)/[||\mathbf{1}_{A_c^r}||_{\boldsymbol{\Sigma}}(t-v+1)]$ and consequently, Eq. (59) is reduced to

$$L_\nu^t(A_c^r) = \frac{u^2(A_c^r)}{2(t-\nu+1)}. \tag{62}$$

Similarly, Eqs. (58) and (60) can be generalized according to a variable radius r to cover at most 50% of the entire area from c.

6.3 Performance Analysis

In this section, we compare the run length performance of the above monitoring statistics using Monte Carlo simulations. In the health surveillance context, we consider monitoring statistics based on only a homogeneous magnitude of shifts in an outbreak cluster, i.e., Eqs. (59) and (62), which assume that the shift magnitude is given or estimated from the data. Both equations will be referred to as regression-adjusted cluster MCUSUM (Rac-MCUSUM) methods. As benchmarks, we will also consider Woodall and Ncube's MCUSUM procedure [112] and Sonesson's approach [94] assuming no spatial correlations. Here, they are referred to as Max-MCUSUM charts with different radii, in which Woodall and Ncube's approach corresponds to $r = 0$, and Sonesson's approach uses a variable radius. For comparison purposes, we will include the radii of 1 and 2 in this study.

Before deploying these methods, we first found the threshold for each surveillance statistic to achieve a target $ARL_0 = 200$. The thresholds for the target ARL_0 are determined based on 100,000 replications. We simulated the ARL_1 values based on 10,000 replications for different shift sizes and change coverage as discussed below.

As illustrated in Figure 20, we consider three types of shift coverage – S-1, S-5, and S-13 – which indicate the 1, 5, and 13 regions of the outbreak areas, respectively. The homogeneous outbreak magnitude is assumed to have 0.5, 1.0, and 2.0 standard deviations for each region. To simplify our discussion, we further assume that any pair of adjacent regions has a correlation coefficient ρ and any two cross-adjacent regions have correlation coefficient $\rho/2$, while any pair of regions with a distance larger than 2 has no correlation. The correlation coefficient ρ is assumed to be 0, 0.2, and 0.5, which represent independence, weak, and medium spatial correlations, respectively. Table 8 presents the out-of-control ARL to trigger. We summarize our finding as follows.

1. *Comparing the Max-MCUSUM and Rac-MCUSUM methods*

   The Rac-MCUSUM method is expected to perform identically to the corresponding Max-MCUSUM chart when no correlations are present, but better when there are correlations. The stronger the correlation is, the more benefits the former method provides. For example, Figure 21 presents the ratio of the ARL's between the Rac-MCUSUM method and the Max-MCUSUM method when an outbreak of S-1 with a magnitude of 1 occurs. It is clear that the advantages of monitoring the regression-adjusted clusters are more significant when the correlation is stronger. The variable radius seems to be more beneficial to the Rac-MCUSUM method when the correlation is weak.

2. *Impact of spatial correlations*

   The important issue of this chapter is to investigate the impact of spatial correlations. For the Max-MCUSUM methods, which were developed assuming no correlations,

it is easy to see that their performance is often adversely affected when correlations are present. The larger the radius is, the stronger the impact will be. Figure 22 shows the percentage of ARL changes from no correlation $\rho = 0$ to the medium correlation of $\rho = 0.5$. For example, to issue a signal for an outbreak of S-1 with magnitude 1.0, the Max-MCUSUM method with $r = 0$, which corresponds to the approach in [112, 100], takes an average of 12.5 steps when no correlation is present but an average of 19.1 steps (about 50% longer) when the correlation is 0.5. Although increasing the radius may make the impact less significant in the outbreak of S-1 case, the Max-MCUSUM method with a variable radius may perform even worse. Its ARL increases from 13.8 to 27.5 (about 100% longer) when the correlation changes from 0 to 0.5. Nevertheless, for outbreaks of patterns S-5 and S-13, increasing the radius may deteriorate the performance of the Max-MCUSUM chart even more significantly.

On the other hand, the impact of spatial correlations on the performance of the regression-adjusted cluster method is always positive, even if the outbreak magnitude is known or estimated from the data. For example, for detecting the same outbreak of S-1 with magnitude 1.0, the ARL of the Rac-MCUSUM method with a given shift magnitude decreases from 12.5 to 1.65 (87% shorter) if $r = 0$ and decreases from 13.9 to 1.76 (87% shorter) if a variable radius is used.

3. *Impact of the radii of the MCUSUM methods*

From Table 8, it is easy to see the importance of matching the radius of the scan statistics to the outbreak coverage. As highlighted in the table, using the same radius as the coverage always results in the smallest ARL regardless of which monitoring method is used. Unfortunately, it is very unlikely that one will know the outbreak coverage in advance. Therefore, it is always helpful to use a variable radius in the scan statistic, which will perform in a manner similar to the optimal fixed radius of

the scan statistics. Tsui *et al.* [103] thoroughly discusses the impact of different radii on the LR-based scan statistics.

When spatial correlations are present, using a variable radius in the Max-MCUSUM method has a considerable impact when the outbreak coverage is small, especially when the correlation is high. For example, in an independent case [103], when an outbreak of S-1 with magnitude 1 occurs, the Max-MCUSUM method with a variable radius takes an average of 13.8 steps to trigger an alert while that with the matched radius ($r = 0$) takes an average of 12.54 steps. When the correlation is 0.5, the Max-MCUSUM chart with the variable radius takes an average of 27.5 steps, which is 44% longer than that with $r = 0$. On the other hand, the adverse impact of the variable radius on the Rac-MCUSUM method is less significant when correlations are present. In the same case of $\rho = 0.5$, the variable radius deteriorates the Rac-MCUSUM method by 7% compared with the matched radius.

4. *When δ is unknown*

When the outbreak magnitude is unknown and a search is conducted using the maximum likelihood principle to estimate the magnitude, both MCUSUM methods will be affected. When the radius and magnitude match those of an outbreak, generally the estimation will deteriorate the performance. However, it is interesting to note that when the radius does not match that of an outbreak, the Rac-MCUSUM method may perform better using the estimated parameters, especially when the correlation is significant. For example, when the correlation is 0.5, the Rac-MCUSUM method with an estimated magnitude always has a shorter ARL when the radius is different from the actual outbreak coverage. This interesting property, brought by the spatial correlations, warrants further investigation.

Table 8: Comparison of ARL for Different Surveillance Methods

Correlation	Magnitude	Coverage	Max-MCUSUM (known magnitude)				Rac-MCUSUM (known magnitude)				Rac-MCUSUM (unknown magnitude)			
			r=0	r=1	r=2	variable	r=0	r=1	r=2	variable	r=0	r=1	r=2	variable
0	0.5	S-1	**35.45**	59.16	67.27	38.05	**35.73**	57.71	66.08	37.71	**43.97**	70.31	75.78	49.14
		S-5	19.95	**9.68**	14.43	10.16	20.02	**9.67**	14.47	10.07	22.44	**11.58**	15.75	12.52
		S-13	15.57	6.18	**4.11**	4.50	15.58	6.19	**4.11**	4.50	15.14	6.95	**4.86**	5.45
	1	S-1	**12.54**	40.31	58.97	13.81	**12.52**	41.62	59.95	13.88	**14.68**	30.37	40.50	16.22
		S-5	7.10	**3.23**	6.46	3.49	7.04	**3.23**	6.61	3.46	7.62	**3.71**	5.34	4.09
		S-13	5.59	2.15	**1.45**	1.62	5.59	2.16	**1.44**	1.61	5.33	2.44	**1.70**	1.88
	2	S-1	**3.99**	30.38	60.76	4.21	**4.04**	30.76	59.87	4.15	**4.58**	10.43	15.37	5.04
		S-5	2.36	**1.17**	2.73	1.26	2.34	**1.17**	2.75	1.26	2.60	**1.32**	1.85	1.43
		S-13	1.94	1.01	**1.00**	1.00	1.95	1.01	**1.00**	1.00	1.95	1.04	**1.00**	1.00
0.2	0.5	S-1	**35.32**	61.81	68.57	49.09	**31.79**	59.40	67.26	35.57	**38.91**	69.15	73.70	44.68
		S-5	20.68	**12.71**	19.27	14.11	24.98	**11.88**	19.67	12.89	30.44	**14.06**	23.19	15.57
		S-13	15.88	8.30	**6.09**	6.40	20.91	10.05	**5.99**	6.68	24.80	12.34	**7.07**	8.13
	1	S-1	**12.54**	43.85	59.88	19.13	**11.09**	41.90	55.73	12.63	**12.67**	30.09	37.29	14.64
		S-5	7.29	**4.25**	8.93	4.85	9.01	**3.89**	9.31	4.36	10.64	**4.44**	7.90	4.98
		S-13	5.71	2.91	**2.12**	2.25	7.75	3.56	**2.03**	2.33	8.90	4.22	**2.36**	2.66
	2	S-1	**3.97**	32.18	55.66	5.34	**3.54**	28.71	46.65	3.78	**4.00**	10.45	13.72	4.55
		S-5	2.44	**1.46**	3.91	1.54	3.02	**1.34**	4.06	1.46	3.56	**1.55**	2.68	1.71
		S-13	1.98	1.09	**1.02**	1.03	2.65	1.25	**1.01**	1.04	3.13	1.53	**1.03**	1.06
0.5	0.5	S-1	**34.58**	67.93	72.71	60.45	**4.77**	7.16	5.91	5.08	**5.23**	5.47	4.97	5.46
		S-5	21.08	**16.07**	24.64	19.71	1.93	**1.32**	1.28	1.49	1.58	**1.40**	1.36	1.54
		S-13	16.33	10.89	**8.46**	9.11	1.80	1.14	**1.08**	1.21	1.34	1.19	**1.12**	1.25
	1	S-1	**19.13**	48.47	62.24	27.54	**1.65**	3.02	2.14	1.76	**1.82**	1.93	1.81	1.93
		S-5	4.85	**5.68**	11.49	6.89	1.00	**1.00**	1.00	1.00	1.00	**1.00**	1.00	1.00
		S-13	2.25	3.88	**2.98**	3.19	1.00	1.00	**1.00**	1.00	1.00	1.00	**1.00**	1.00
	2	S-1	**3.96**	33.98	54.25	8.05	**1.00**	1.37	1.29	1.00	**1.00**	1.01	1.00	1.00
		S-5	2.55	**1.89**	5.13	2.17	1.00	**1.00**	1.00	1.00	1.00	**1.00**	1.00	1.00
		S-13	2.05	1.34	**1.15**	1.19	1.00	1.00	**1.00**	1.00	1.00	1.00	**1.00**	1.00

6.4 Example - Breast Cancer in New Hampshire

The example in this chapter is based on female breast cancer data in New Hampshire during 1968-1994, which has been studied by Rogerson and Yamada [88]. The annual data set was obtained from the Compressed Mortality File from the CDC. We investigated the mortality of female breast cancer for ten counties in New Hampshire. The map of New Hampshire is shown in the left panel of Figure 23, which was obtained from the U.S. Census Bureau (http://quickfacts.census.gov/qfd/index.html).

In this example, in order to model the number of breast cancer counts using a normal distribution, we transformed the discrete data using the square root transformation introduced in [3]. Assuming y follows a Poisson distribution with parameter λ, Anscombe [3] showed that $x = 2\sqrt{y + 3/8}$ approximately follows a normal distribution, which has mean $2\sqrt{\lambda}$ and variance 1 if λ is greater than 20. For spatiotemporal surveillance, since the population is often non-homogeneous in both temporal and spatial respects, we introduced the following transformation to account for the heterogeneity of populations. For each county c at time t, let y_{ct} be the observed mortality count, N_{ct} be the population size, $\bar{N}_c$ be the average population of country c under the time window we considered. The transformed data are obtained as $x_{ct} = 2\sqrt{y_{ct}\bar{N}_c/N_{ct} + 3/8}$. Finally, $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{pt})'$ approximately follows a multivariate normal distribution $N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma})$. Note that x_{ct} has approximately constant variance that does not change even when the mortality rates increase.

Assume we are interested in detecting a special increase of the breast cancer rate as early as possible, e.g., three times the standard deviation in any region. The formulation of the problem follows,

$$H_0 : \mu_{ct} = \mu_{c0} \text{ for all regions } c \text{ and time } t.$$

against

$$\begin{aligned} H_1 : \quad & \mu_{ct} = \mu_{c0} \text{ for some regions } c, \text{ when time } t = 1, 2, ..., \nu - 1 \\ & \mu_{ct} = \mu_{c0} + 3\sigma_c \text{ for some regions } c, \text{ when time } t = \nu, \nu + 1, ... \end{aligned}$$

where σ_c is the standard deviation at region c, and ν is the unknown time of any outbreak.

Figure 2 presents the mortality rate per 10,000 in New Hampshire from 1968 to 1994. The first step is the estimation of a baseline rate for each region. Since the mortality rate appears to be quite stable between 1968 to 1982, we used this period of mortality count data to estimate the baseline rate in each county. If the mean of the Poisson data is known, we can calculate the mean of the transformed data directly from Anscombe's result in [3]. However, the mean of mortality counts is unknown in practice. We then used the transformed data to estimate the mean between 1968 to 1982. For the covariance matrix, we only needed to estimate the correlations among different regions. We used three approaches in this chapter. In the first approach, we assumed that any two regions were independent, so the covariance matrix was the identity matrix. In the second approach, we simply calculated an unrestricted covariance matrix from the transformed data. In the third approach, we used Daniels and Kass's method, which shrinks the eigenvalues of the covariance matrix by a simple hierarchical model [20]. For the first approach, we used the Max-MCUSUM statistic for detection. For the second and third approaches, we used the Rac-MCUSUM method, which takes into account spatial correlations. These two statistics are denoted as Rac-MCUSUM(U) and Rac-MCUSUM(E), respectively.

After we estimated the baseline rates and covariance matrix, we applied the above surveillance methods to quickly detect emerging clusters that have increasing mortality rates. Because the map of New Hampshire shows an irregular pattern, we had to adjust the definition of clusters and the location of centers in counties to X-Y coordinates, as shown in Figure 23 (b). We defined the cluster around each center in the X-Y coordinates as those centers that have distance less than radius r. In this chapter, we considered (i) $r = 0$ and (ii) $r = 5$. In order to compare the detection methods, we set $ARL_0 = 100$. Since each method has a different scale of the control limit, Figures 24 and 25 plot the monitoring statistics re-scaled by their corresponding control limits for each method. An alarm is triggered when the scaled statistic exceeds 1.

As shown in Figure 24 when $r = 0$, based on the independence assumption, Max-MCUSUM triggered an alarm in 1985. On the other hand, Rac-MCUSUM with an unrestricted covariance estimate triggered an alarm in 1983 since it utilizes the spatial correlation information in the surveillance. The scaled value of the Rac-MCUSUM statistics was much higher than 3 in 1983, so it is not shown in the plot. In addition, Rac-MCUSUM based on Daniels and Kass's covariance estimator also triggered an alarm in 1983. Similarly, when $r = 5$, as shown in Figure 25, all three methods triggered alarms in 1983. However, the Max-MCUSUM statistic in 1983 is almost 1, so the signal is not significant compared to the signals of the Rac-MCUSUM methods. Nevertheless, the signal becomes more significant after 1985. Therefore, Rac-MCUSUM was more sensitive to detecting the mortality increase than Max-MCUSUM.

In addition to the detection delay, we also investigated the clusters most strongly contributing to the alarm. As shown in Figure 26, all methods identify the emerging cluster as Coos County, which is consistent with [88]. In this example, since the covariance matrix was estimated based on the data from a short time period, cluster identification may be questionable. In practice, as pointed out in [94], possible clusters need to be investigated by epidemiologists.

6.5 Conclusions

Based on the UIT and GLRT principles (Jiang and Tsui [44]), we develop a general framework for multivariate change-point detection problems to interpret and relate existing multivariate CUSUM procedures for temporal surveillance. By taking advantage of the cluster characteristics, we propose a MCUSUM method using regression-adjusted clusters for spatiotemporal surveillance in the presence of spatial correlations. Different monitoring statistics have been developed under various assumptions. The proposed monitoring statistics are shown to outperform the applications of existing MCUSUM charts in public health

surveillance, especially when the spatial correlations are significant. The case of an unknown outbreak magnitude is also investigated. A numerical example also illustrates the potential effectiveness of the proposed method.

An important finding of this research is the importance of matching the radius and magnitude of a scan statistic with those of an actual outbreak. Currently, most spatiotemporal surveillance methods are based on the assumption of a known/given outbreak pattern. Our research suggests that estimating the outbreak magnitude may help the detection of outbreaks when there are medium-sized spatial correlations. It would be important to estimate both the outbreak coverage and the magnitude when we have little knowledge about them and substitute the estimates into the corresponding scan statistic in practice. This will be further pursued in future research.

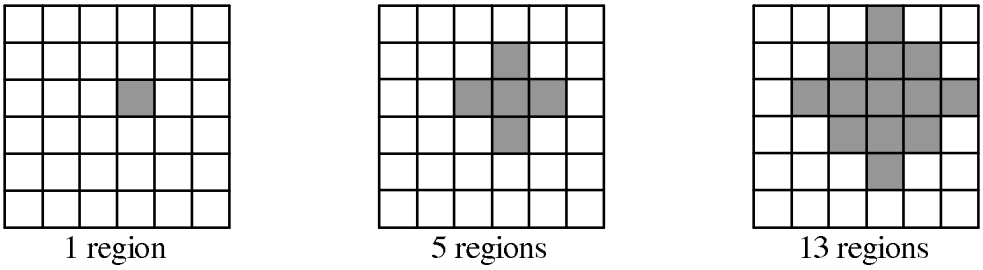


Figure 20: Three symmetric outbreak patterns (S-1, S-5, and S-13)

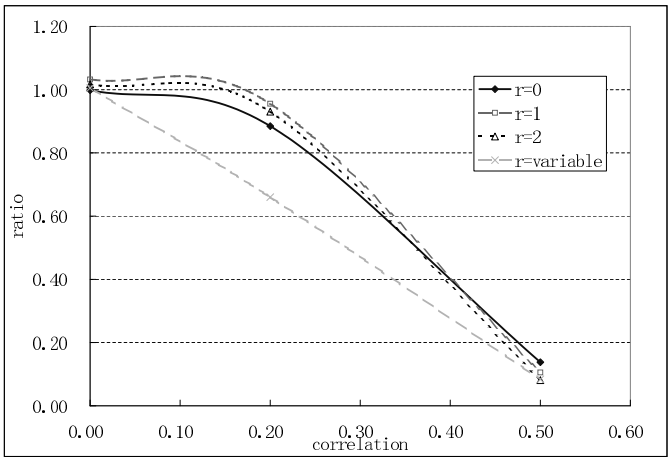
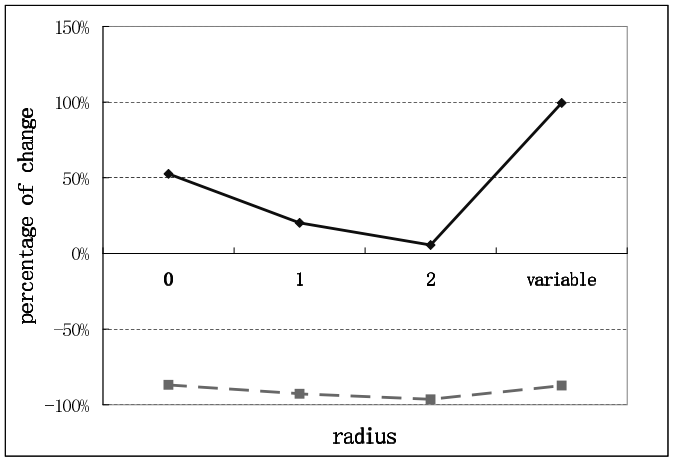
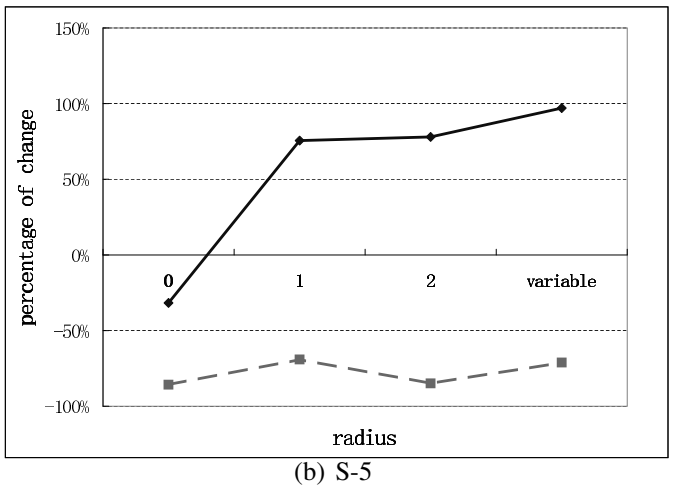


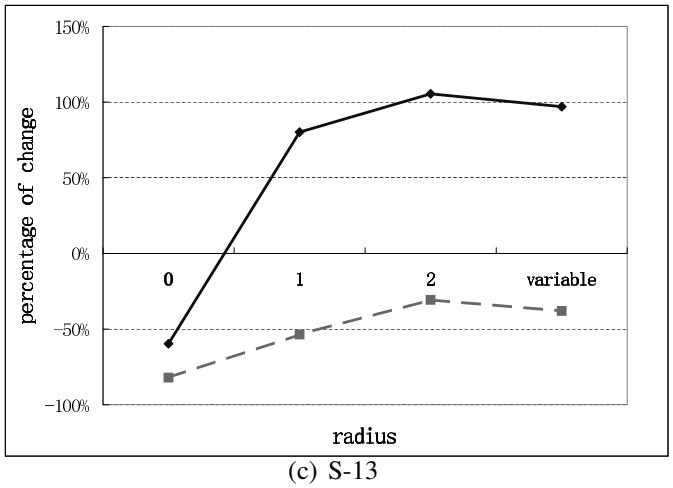
Figure 21: The ARL comparison of Rac-MCUSUM and Max-MCUSUM methods when correlations are present (outbreak S-1 and $\delta = 1$)



(a) S-1

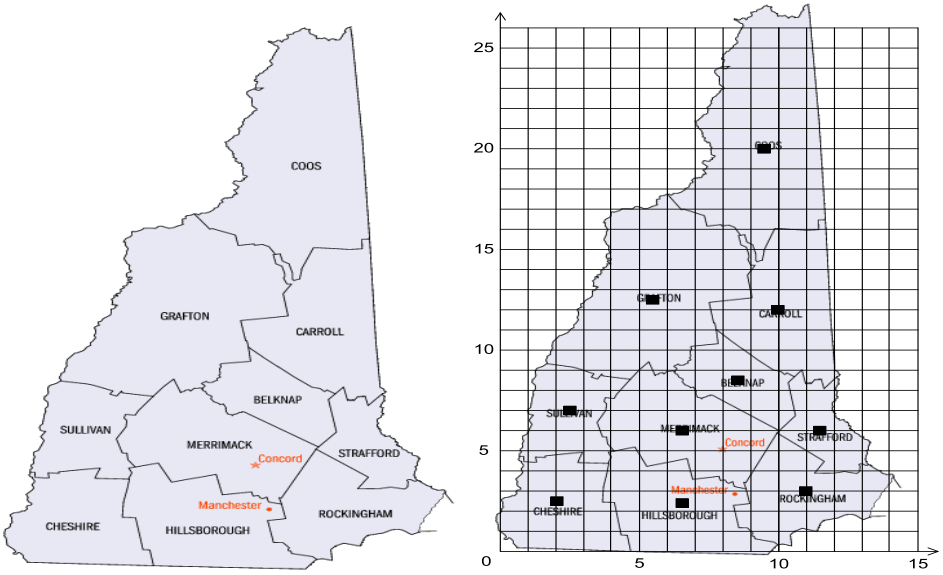


(b) S-5



(c) S-13

Figure 22: The impact of spatial correlations on different surveillance methods ($\rho = 0.5$ and $\delta = 1$) (solid line: Max-MCUSUM, dashed line: Rac-MCUSUM with a known magnitude)



(a) The map of New Hampshire (b) X-Y coordinate

Figure 23: The map of New Hampshire and the X-Y coordinate

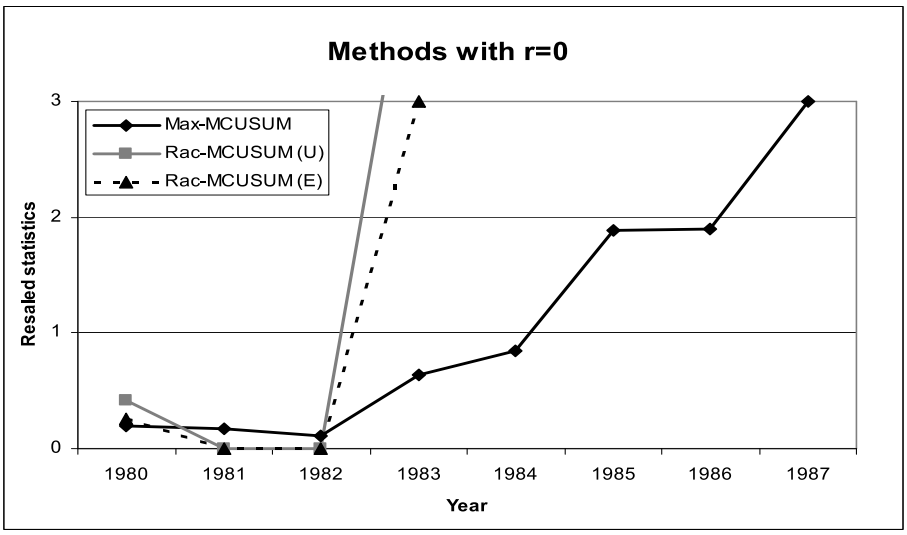


Figure 24: Surveillance statistics with $r = 0$ for breast cancer data in New Hampshire

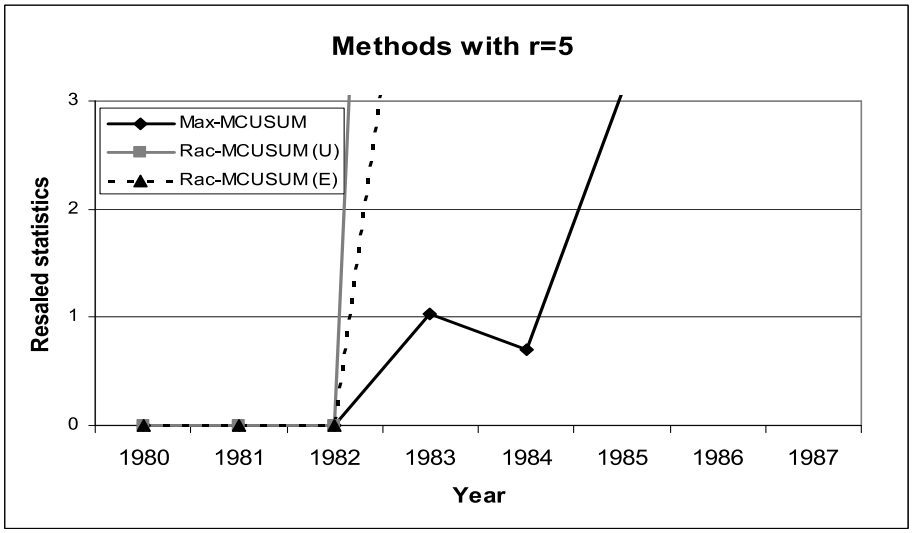


Figure 25: Surveillance statistics with $r = 5$ for breast cancer data in New Hampshire

Figure 26: The cluster detected by the methods

REFERENCES

[1] Aalen OO. Nonparametric inference for a family of counting processes. *Annals of Statistics* 1978; 6: 701-726.

[2] Alwan LC, Roberts HV. Time-series modeling for statistical process-control. *Journal of Business & Economic Statistics* 1988; 6:87-95.

[3] Anscombe FJ. The transformation of Poisson, binomial, and negative binomial data. *Biometrika* 1948; 35:246-254.

[4] Arellano M. *Panel data econometrics* Oxford University Press, 2003.

[5] Balakrishnan N, Koutras MV. *Runs and scans with applications.* Wiley: New York, 2002.

[6] Baron M, Tartakovsky AG. Asymptotic optimality of change-point detection schemes in general continuous-time models. *Sequential Analysis* 2006; 25: 257–296.

[7] Basseville M, Nikiforov lV. *Detection of abrupt changes - theory and application.* Prentice-Hall: Englewoood Cliffs, N.J., 1993.

[8] Benneyan JC. Statistical quality control methods in infection control and hospital epidemiology, part I: Introduction and basic theory. *Infection Control and Hospital Epidemiology* 1998a; 19: 194–214.

[9] Benneyan JC. Statistical quality control methods in infection control and hospital epidemiology, part II: Chart use, statistical properties, and research issues. *Infection Control and Hospital Epidemiology* 1998b; 19: 265–283.

[10] Borror CM, Champ CW, Rigdon SE. Poisson EWMA control chart. *Journal of Quality Technology* 1998; 30:352-361.

[11] Borror CM, Rigdon SE, Champ CW. An exponentially weighted moving average control chart for attribute data. *Proceedings of the 23rd Annual Modeling and Simulation Conference* Pittsburgh, PA. 1775-1782.

[12] Brook D, Evans DA. An approach to the probability distribution of CUSUM run length. *Biometrika* 1972; 59:539-549.

[13] Burkom HS, Murphy SP, Shmueli G. Automated time series forecasting for biosurveillance. *Statistics in Medicine* 2007; 26:4202-4218.

[14] Chan HP, Lai TL. Saddlepoint approximations and nonlinear boundary crossing probability of Markov random walks. *The Annals of Applied Probability* 2003; 13:395-429.

[15] Chen C. A surveillance system for congenital malformations. *Journal of the American Statistical Association* 1978; 73:323-327.

[16] Crosier RB. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics* 1988; 30:291-303.

[17] Crowder SV. A simple method for studying run length distributions of exponentially weighted moving average control charts. *Technometrics* 1987; 29:401-407.

[18] Crowder SV. Design of exponentially weighted moving average schemes. *Journal of Quality Technology* 1989; 21:155-162.

[19] Crowder SV, Hamilton M. Average run lengths of EWMA controls for monitoring a process standard deviation. *Journal of Quality Technology* 1992; 24:12-21.

[20] Daniels MJ, Kass RE. Shrinkage estimators for covariance matrices. *Biometrics* 2001; 57:117-1184.

[21] Fan J, Gijbels I. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of Royal Statisticla Society Ser.B* 1995; 57: 371-394.

[22] Fienberg SE, Shmueli, G. Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statistics in Medicine* 2005; 24:513-529.

[23] Fraker SE, Woodall WH, Burkom HS. A Note on the Poisson Likelihood Ratio Test Statistic for Kulldorff's Scan Methods. *Communications in Statistics - Theory and Methods* 2008; 37:998-1001.

[24] Fricker RD, Jr. Syndromic surveillance. *Encyclopedia of Quantitative Risk Assessment* Melnick E and Everitt B.(eds.), John Wiley & Sons Ltd., 2008; 1743-1752.

[25] Fricker RD, Jr., Rolka HR. Protecting against biological terrorism: statistical issues in electronic biosurveillance. *Chance.* 2006; 19:4-13.

[26] Frisén M. Properties and use of the Shewhart method and its followers. *Sequential Analysis* 2007; 26:171-193.

[27] Gan FF. Monitoring Poisson observations using modified exponentially weighted moving average control charts. *Communications in Statistics – Simulation and Computation* 1990; 19:103-124.

[28] Gijbels I, Pope A, Wand MP. Understanding exponential smoothing via kernel regression. *Journal of Royal Statistical Society, Ser. B* 1999; 61:39-50.

[29] Glaz J, Balakrishnan N (eds). *Scan statistics and applications.* Birkhäuser: Boston, MA, 1999.

[30] Glaz J, Naus J, Wallenstein S. *Scan statistics.* Springer: New York, 2001.

[31] Grigg O, Farewell V. An overview of risk-adjusted charts. *Journal of the Royal Statistical Society, Series A* 2004; 167:523-539.

[32] Han SW, Tsui K-L, Ariyajunyab B, Kim SB. A comparison of CUSUM, EWMA, and temporal scan statistics for detection of increases in Poisson rates. *Quality and Reliability Engineering International* 2009; DOI: 10.1002/qre.1056.

[33] Hawkins DM. Multivariate quality control based on regression-adjusted variables. *Technometrics* 1991; 33(1):61-75.

[34] Hawkins DM. Regression adjustment for variables in multivariate quality control. *Journal of Quality Technology* 1993; 25(3):170-182.

[35] Hawkins DM, Olwell DH. *Cumulative sum charts and charting for quality improvement.* Springer Verlag: New York, 1998.

[36] Healy JD. A note on multivariate CUSUM procedures. *Technometrics* 1987; 29(4):409-412.

[37] Hill GB, Spicer CC, Weatherall JAC. The computer surveillance of congenital malformations. *BMJ.* 1968; 24:215-218.

[38] Holt CC. *Forecasting seasonals and trends by exponentially weighted moving averages.* ONR Research Memorandum, Carnegie Institute 1957; 52.

[39] Hunter JS. The exponentially weighted moving average. *Journal of Quality Technology* 1986; 18:203-210.

[40] Hutwagner L, Thompson GM, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). *Journal of Urban Health* 2003; 80:i89-96.

[41] Ismail NA, Pettitt AN, Webster RA. Online monitoring and retrospective analysis of hospital outcomes based on a scan statistic. *Statistics in Medicine* 2003; 22:2861-2876.

[42] Jiang W, Au T, Tsui KL. A statistical process control approach for customer activity monitoring. *IIE Transactions* 2007; 39:235-249.

[43] Jiang W, Han SW, Tsui K-L, and Woodall W. Spatiotemporal surveillance methods in the presence of spatial correlation. Accepted to *Statistics in Medicine* 2010.

[44] Jiang W, Tsui KL. A theoretical framework and effciency study of multivariate statistical process control charts. *IIE Transactions* 2008; 40:650-663.

[45] Jiang W, Wu HQ, Tsung FG, Nair VN, Tsui K-L. Proportional integral derivative charts for process monitoring. *Technometrics* 2002; 44:205-214.

[46] Joner MD, Jr., Woodall WH, Reynolds MR, Jr. Detecting a rate increase using a Bernoulli scan statistic. *Statistics in Medicine* 2008; 27:2555-2575.

[47] Joner MD, Jr., Woodall WH, Reynolds MR, Jr., Fricker RD. A one-sided MEWMA chart for health-related surveillance. *Quality and Reliability Engineering International* 2008; 24:503-518

[48] Kleinman K, Abrams A, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. *Epidemiology and Infection* 2005; 133:409-419.

[49] Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology* 2004; 159:217-224.

[50] Knox G. The detection of space-time interactions. *Applied Statistics* 1964; 13:25-29.

[51] Kulldorff M. A spatial scan statistic. *Communications in Statistics - Theory and Methods* 1997; 26:1481-1496.

[52] Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society, Series A* 2001; 164(1):61-72.

[53] Kulldorff M, Zhang Z, Hartman J, Heffernan R, Huang L, Mostashari F. Benchmark data and power calculations for evaluating disease outbreak detection methods. *Morbidity and Mortality Weekly Report* 2004; 53:144-151.

[54] Lai TL. Sequential change-point detection in quality control and dynamical systems (with discussions). *Journal of Royal Statistical Society, Series B* 1995; 57:613-658.

[55] Lai TL. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory* 1998; 44:2917-2929.

[56] Lai TL. Sequential analysis: some classical problems and new challenges. *Statistica Sinica* 2001; 11: 303-408.

[57] Lorden G. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics.* 1971; 42:1897-1908.

[58] Lowry CA, Woodall WH, Champ CW, Rigdon SE. A multivariate exponentially weighted moving average control chart. *Technometrics* 1992; 34:46-53.

[59] Lucas JM. Counted data CUSUMs. *Technometrics* 1985; 27:129-144.

[60] Lucas JM, Saccucci MS. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics* 1990; 32:1-12.

[61] Marshall JB, Spitzner DJ, Woodall WH. Use of the local Knox statistic for the prospective monitoring of disease occurrences in space and time. *Statistics in Medicine* 2007; 26:1579-1593.

[62] Mei Y. Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* 2010; 97:419-433.

[63] Mei Y, Han SW, Tsui K-L. Early detection of a change in Poisson rate after accounting for population size effects. *Statistica Sinica* 2010, in press.

[64] Montgomery DC. *Statistical quality control, 5th edition*. John Wiley & Sons: New York, 2005.

[65] Montgomery DC, Mastrangelo CM. Some statistical process-control methods for autocorrelated data. *Journal of Quality Technology* 1991; 23:179-193.

[66] Morrison DF. *Multivariate statistical methods, 2nd de.* McGraw-Hill Book Company: New York, NY, 1976.

[67] Moustakides GV. Optimal stopping times for detecting changes in distribution. *Annals of Statistics* 1986; 14:1379-1387.

[68] Naus J, Wallenstein. Temporal surveillance using scan statistics. *Statistics in Medicine* 2006; 25(2):311-324.

[69] Ng CH, Case KE. Development and evaluation of control charts using exponentially weighted moving averages. *Journal of Quality Technology.* 1989; 21:242-250.

[70] Page ES. Continuous inspection schemes. *Biometrika.* 1954; 41: 100-115.

[71] Peskir G, Shiryaev A. *Optimal stopping and free-boundary problems.* Birkhäuser Verlag, Basel, Switzerland, 2006

[72] Pignatiello JJ, Jr., Runger GC. Comparisons of multivariate CUSUM charts. *Journal of Quality Technology* 1990; 22:173-186.

[73] Pinheiro JC, Douglas MB. *Mixed-effects models in S and S-PLUS*, New York: Springer, 2000.

[74] Pollak M. Optimal detection of a change in distribution. *Annals of Statistics.* 1985; 13: 206-227.

[75] Pollak M, Siegmund D. A diffusion process and its applications to detecting a change in the drift of a Brownian motion. *Biometrika* 1985; 72: 267-280.

[76] Pollak M, Siegmund D. Sequential detection of a change in a normal mean when the initial value is unknown. *Annals of Statistics* 1991; 19: 394-416.

[77] Poor HV, Hadjiliadis O. *Quickest detection.* Cambridge University Press, New York, 2008.

[78] Raubertas RF. An analysis of disease surveillance data that uses the geographic locations of the reporting units. *Statistics in Medicine* 1989; 8:267-271.

[79] Resnick SI. *A probability path.* Birkhäuser: Boston, MA, 2001.

[80] Reynolds MR, Jr., Stoumbos ZG. A CUSUM chart for monitoring a proportion when inspecting continuously. *Journal of Quality Technology* 1999; 31:87-108.

[81] Ritov Y. Decision theoretic optimality of the CUSUM procedure. *Annals of Statistics* 1990; 18: 1464-1469.

[82] Roberts SW. Control chart tests based on geometric moving averages. *Technometrics* 1959; 1:239-250.

[83] Roberts SW. A comparison of some control chart procedures. *Technometrics* 1966;8: 411-430.

[84] Robinson PB, Ho TY. Average run lengths of geometric moving average charts by numerical methods. *Technometrics.* 1978; 20:85-93.

[85] Rogerson PA. Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine* 1997; 16(18):2081-93.

[86] Rogerson PA. Monitoring point patterns for the development of space-time clusters. *Journal of the Royal Statistical Society, Series A* 2001; 164:87-96.

[87] Rogerson PA, Yamada I. Approaches to syndromic surveillance when data consist of small regional counts. *Morbidity and Mortality Weekly Report* September 24, 2004; 53(Sp):79-85.

[88] Rogerson PA, Yamada I. Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches. *Statistics in Medcine* 2004; **23**:2195-2214.

[89] Rolka H, Burkom H, Cooper GF, Kulldorff M, Madigan D, Wong W-K. Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: Research needs. *Statistics in Medicine.* 2007; 26:1834-1856.

[90] Ryan AG, Woodall WH. Control Charts for Poisson Count Data with Varying Sample Sizes. *Journal of Quality Technology* 2010, in press.

[91] Shiryaev AN. On optimum methods in quickest detection problems. *Theory of Probability and Its Applications.* 1963; 8: 22-46.

[92] Shmueli G, Burkom H. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics.* 2010; 52: 39-51.

[93] Shu L, Jiang W, Wu S. A one-sided EWMA control chart for monitoring process means. *Communications in Statistics - Simulation and Computation.* 2007; 36:901-920.

[94] Sonesson C. A CUSUM framework for detection of space-time disease clustering using scan statistics. *Statistics in Medicine* 2007; 26:4770-4789.

[95] Sparks R, Carter C, Graham PL, Muscatello D, Churches T, Kaldor J, Turner R, Zheng W, Ryan L. Understanding sources of variation in syndromic surveillance for early warning of natural or intentional disease outbreaks. *submitted for publication* 2009.

[96] Srivastava MS, Wu YH. Comparison of EWMA, CUSUM and Shiryaev-Roberts procedures for detecting a shift in the mean. *Annals of Statistics* 1993; 21:645-670.

[97] Srivastava MS, Wu YH. Evalution of optimum weights and average run lengths in EWMA control schemes. *Comm. Statist. Theory Method.* 1997; 26:1253-1267.

[98] Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000; 1:441-452.

[99] Tango T. A class of tests for detecting general and focused clustering of rare diseases. *Statistics in Medicine* 1995; 14(21-22):2323-2334.

[100] Tartakovsky AG, Veeravalli VV. Change-point detection in multichannel and distributed systems with applications *Applications of Sequential Methodologies* Mukhopadhyay N, Datta S, Chattopadhyay S, editors. Marcel Dekker, Inc. New York. 2004; 331-363.

[101] Testik MC, Runger GC. Multivariate one-sided control charts. *IIE Transactions* 2006; 38:635-645.

[102] Tsui K-L, Chiu W, Gierlich P, Goldsman D, Liu X, Maschek T. A review of healthcare, public health, and syndromic surveillance. *Quality Engineering* 2008; 20:435-450.

[103] Tsui K-L, Han SW, Jiang W, Woodall WH. CUSUM and likelihood ratio tests for spatial and spatiotemporal surveillance. Submitted

[104] Tsui K-L, Woodall WH. Multivariate control charts based on loss functions. *Sequential Analysis* 1993; 12:79-92.

[105] Wald A. *Sequential Analysis*, New York: John Wiley and Sons, 1947.

[106] Weatherall JAC, Haskey JC. Surveillance of malformations. *BMJ* 1976; 32:39-44.

[107] White CH, Keats JB. ARLs and higher order run length moments for Poisson CUSUM. *Journal of Quality Technology* 1996; 28:363-369.

[108] Winters PR. Forecasting sales by exponentially weighted moving averages. *Management Science* 1960; 6:324-342.

[109] Woodall WH. Control Charting Based on Attribute Data: Bibliography and Review. *Journal of Quality Technology* 1997; 29:172-183.

[110] Woodall WH. The use of control charts in health-care and public health surveillance. *Journal of Quality Technology* 2006; 38:89-104.

[111] Woodall WH, Marshall JB, Joner MD, Jr., Fraker SE, Abdel-Salam G. On the use and evaluation of prospective scan methods for health-related surveillance. *Journal of the Royal Statistical Society. Series A* 2008; 171:223-237.

[112] Woodall WH, Ncube MM. Mutivariate CUSUM quality control procedures. *Technometrics* 1985; 38(3):291-303.

[113] Yakir B. Optimal detection of a change in distribution when the observations form a Markov Chain with a finite state space. Eds E. Carlsten, H-G. Muller, D. Siegmund. *Change-Point Problems, IMS Lecture Notes Monograph Series*, 1994; 23:346-358.

[114] Yao Q. Asymptotically optimal detection of a change in a linear model. *Sequential Analysis* 1993; 12:201-210.

[115] Yashchin E. Weighted cumulative sum technique. *Technometrics* 1989; 31:321-338.

[116] Yashchin E. Statistical control schemes: Methods, applications and generalizations. *Int. Statist. Rev.* 1993; 61:41-66.

VITA

Sung Won Han received his undergraduate degree in Industrial System and Information Engineering from Korea University in August 2003. He joined the master's program in the School of Industrial and Systems Engineering at Georgia Tech. He received an M.S. Degree in Operations Research in December 2006, an M.S in Statistics in December 2007, and an M.S in Mathematics in May 2010. He also started a Ph.D. program in January 2005 in the same department. His concentration is change point problems–temporal and spatiotemporal cases under non-homogenous baseline or correlated time series. Main applications include healthcare/disease surveillance.